PARTICLE ASTROPHYSICS

The early universe and cosmic structures

edited by

J.-M. Alimi, A. Blanchard, A. Bouquet,
F. Martin de Volnay and J. Tran Thanh Van

microcosm CERN

Editions Frontières

Les Arcs 1990

CONTENTS

IV. DARK MATTER AND WIMPS

V. LEP (ASTRO) PHYSICS

VI. MASS AND LIGHT INCOSMIC STRUCTURES

PRIMEVAL UNIVERSE

INITIAL CONDITIONS FOR INFLATION

Tsvi Piran and Dalia S. Goldwirth
The Racah Institute for Physics,
The Hebrew University, Jerusalem, ISRAEL

ABSTRACT

We discuss the required initial conditions for inflation. We show that a generic kinetic term does not prevent the onset of inflation but that the gradient term might. Domains larger than the horizon in which $\Phi >$ (a few) $\times M_{pl}$ are required for the onset of inflation. If we assume that the kinetic, gradient and potential terms are of the same order large domains with $\Phi >$ (a few) $\times M_{pl}$ exist and inflation can begin. However if we assume that the scalar field is in a thermal equilibrium, the potential term must be smaller and such domains will not be present and there is no room for inflation. It seems that inflation does not free us completely from the need for special initial conditions. However, Linde [1] has pointed out that inflation can be eternal. He stresses the fact that inflation will never cease, but this also means that it did not necessarily have a beginning. We argue that this is the simplest solution to the initial conditions for the universe and that inflation might replace the "traditional" singularity of the Big Bang model.

The standard big bang model faces two well known problems - the horizon problem and the flatness problem. Both problems can be considered not to be problems if we accept the idea of very special initial conditions. We can simply declare that there is no problem - God created a Friedmann Universe with extremely small initial curvature and acausal homogeneity. Dicke and Peebles [2] have pointed out that to a physicist this should seem quite unnatural. The inflationary paradigm [3] suggests that a physical mechanism, inflation, rather than ad hoc initial conditions has lead to the observed Universe.

For inflation we need a slowly varying scalar field ($\dot{\Phi}/\Phi < \dot{R}/R$), whose potential dominates the energy density of the Universe:

$$\rho_{total} \approx \rho_\Phi = \frac{\dot{\Phi}^2}{2} + \frac{(\nabla\Phi)^2}{2R^2} + V(\Phi) \approx V(\Phi) \quad . \tag{1}$$

The potential acts as an effective cosmological constant giving rise to a de Sitter phase - commonly called inflation. During this phase the scale factor of the Universe, R, increases exponentially, making the curvature term negligibly small and the horizon exponentially large.

There are two main variants of inflation: "chaotic inflation" [4], with $V(\Phi) = \lambda\Phi^n$, and "new inflation" [5]. In the sequel we focus on "chaotic inflation" since the problems that we raise are only exasperated in the "new inflation" scenario. For simplicity we use an $n = 2$ potential, for which $\lambda = m^2/2$ (m is the mass of the scalar field). None of our arguments depends on this specific choice.

Inflation solves both the horizon and the flatness problems - but does the inflationary paradigm live to its promise to free cosmology from the worry about initial conditions? We know that inflation takes place when initially the potential dominates the energy density $\rho_{total} \approx \rho_\Phi \approx V$ (and when $\Phi_0 > $ (a few) $\times M_{pl}$). In general the kinetic term $\dot{\Phi}^2/2$ and the gradient term $(\nabla\Phi)^2/2R^2$ also contribute to ρ_Φ. The central question is: what happens under initial conditions when these other terms are comparable or larger than the potential term?

The effect of the kinetic term has been discussed quite extensively [6,7,8]. By now it is well known that "Chaotic inflation" develops even when the kinetic term dominates initially the energy density provided that $\Phi_0 > $ (a few) $\times M_{pl}$. To see this consider the Friedmann equation:

$$H^2 + \frac{k}{R^2} = \frac{8\pi}{3M_{pl}^2}\rho \tag{2}$$

and the homogenous scalar field equation:

$$\ddot{\Phi} + 3H\dot{\Phi} + V'(\Phi) = 0 \quad . \tag{3}$$

When the kinetic term dominates $\rho \approx \dot{\Phi}^2/2$ and if we ignore the curvature term in the Friedmann equation we find:

$$\ddot{\Phi} \approx -\frac{\sqrt{12\pi}}{M_{pl}}\dot{\Phi}^2 \tag{4}$$

In this case

$$\dot{\Phi} \propto 1/t \quad \text{and} \quad \Phi \propto \log t \quad . \tag{5}$$

$\dot{\Phi}$ decreases rapidly while Φ remains almost constant - the kinetic term disappears while the potential term remains almost unchanged, and inflation begins. It is important, for latter usage to recall that during the phase when the kinetic term dominates the energy density, $p = \rho$, the scale factor, R, varies like $t^{1/3}$ and the density decreases like
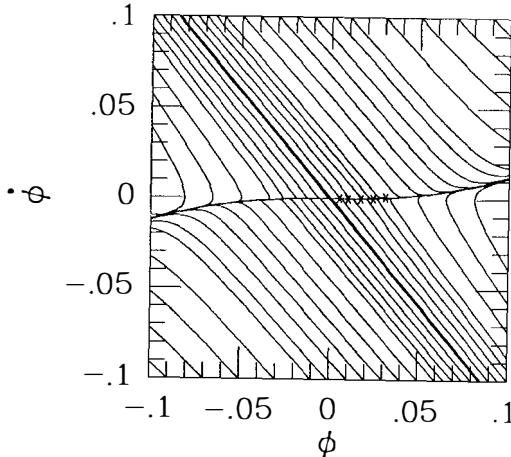
Fig. 1: *Trajectories in the $(\Phi, \dot{\Phi})$ plane for "Chaotic inflation". The horizontal curve with $\dot{\Phi} \approx 0$ corresponds to the slow rolling phase and to inflation.*

R^{-6}. The rapid decrease of $\dot{\Phi}$ can be seen easily from the trajectory of the solution in the $(\dot{\Phi}, \Phi)$ plane (Fig. 1). The slow rolling phase, where

$$\dot{\Phi} \approx -\frac{V'}{3H} \quad , \tag{6}$$

is an attractor and the trajectories converge to it quickly. For "chaotic inflation" this is the inflationary phase and hence "chaotic" inflation is generic (as far as a kinetic term is concerned).

Recently one of us [9] has shown that in "new inflation" the slow rolling phase is also an attractor (see Fig 2), but only a small fraction of this line (a tiny region where $\Phi < 0.01\sigma$) corresponds to inflation. Hence a kinetic term might prevent the onset of new inflation even if $\Phi \approx 0$ as required for new inflation.

The previous analysis was done in the context of a Universe in which the curvature term can be neglected. Belinski et al [8] have pointed out that in a closed universe it is possible that the universe will collapse before inflation starts. Recall that the curvature term decrease like R^{-2}, the kinetic term like R^{-6} and the potential term remains roughly constant. The fact that the density decreases faster then the curvature causes collapse in regular Friedmann cosmology. Collapse can happen and prevent inflation from starting if the energy density that is in the kinetic term does not become negligible before the turnover condition:

$$\frac{1}{R^2} = \frac{8\pi}{3M_{pl}^2}\rho \quad , \tag{7}$$

is satisfied. The energy density can be related to the initial values R_0 and Φ_0 using the scaling laws that we have mentioned earlier:

$$\rho = \frac{\dot{\Phi}_0^2}{2}\frac{R_0^6}{R^6} + V(\Phi_0) \quad , \tag{8}$$
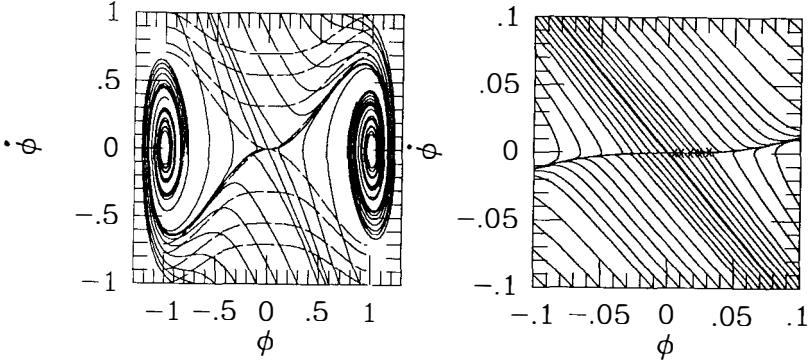
Fig. 2: *Trajectories in the $(\Phi,\dot{\Phi})$ plane for "new inflation". The axis are scaled so that the minimum of the potential is at $\Phi = 1$. The dashed lines are $\rho = \Pi^2/2 + V(\Phi) = $ const lines (from $\rho = V(0)$ to $\rho = 2V(0)$). Fig. 2b is an enlargement of the region around the origin of Fig. 2a. The $*$ corresponds to points that have N (number of e-folds) of 21600, 100, 60, 10, and 1. It can be seen that N drops very quickly, and only the first three curves around zero lead to sufficient inflation.*

between R and ρ and the initial values R_0 and ρ_0. Recollapse begins when Eq. 7 is satisfied and this can happen provided that the energy density is dominated at this stage by the kinetic term. This lead to the following condition:

$$\frac{M_{pl}^3}{(8\pi/3)^3 V^2(\Phi_0) R_0^6} < \frac{\dot{\Phi}_0^2}{2} \quad , \tag{9}$$

which has to be satisfied in order that the universe will not recollapse before it starts inflating. Similar condition was found numerically by Belinski et al. [8].

Until recently the onset of inflation was tested only under homogeneous initial conditions and it was not clear what are the possible consequences of a gradient term. However, to justify its claim to fame, inflation should be able to emerge from truly generic inhomogeneous initial conditions. One can [7] approximate the gradient term by an effective homogenous term $\delta\Phi^2/\Delta^2 R^2$ in the energy density where $\delta\Phi$ is the magnitude of the variations in Φ and Δ a typical comoving wavelength of these variations. If the inhomogeneous perturbations decouple from the evolution of the average scalar field, $\bar{\Phi}$, they evolve according to the approximate equation:

$$\bar{\delta}\Phi + 3H\dot{\delta}\Phi + \frac{\delta\Phi}{\Delta^2 R^2} = 0 \quad . \tag{10}$$

and H can be approximated by:

$$H^2 \approx \frac{8\pi}{3M_{pl}^2} \frac{\delta\Phi^2}{2\Delta^2 R^2} \quad . \tag{11}$$

One can easily show that $\delta\Phi$ decays like R^{-1} and the energy density of the gradients decays like a radiation field, i.e. $\rho_{grad} \propto R^{-4}$. This decay is very rapid and Φ is almost unchanged in the time that $\delta\Phi$ decays. Brandenberger et al. [10] use a more complicated approximation in which they solve the inhomogeneous wave equation combined with a homogenous Friedmann equation (for which they use the average energy density as a source) and obtain similar results. Like with the kinetic term one has to worry, in the case of a closed universe, whether the universe will not recollapse before inflation starts. Using the same approach we derive the condition

$$\frac{M_{pl}^4}{V(\Phi_0)R_0^4} < \frac{\delta\Phi^2}{2\Delta^2 R_0^2} \quad . \tag{12}$$
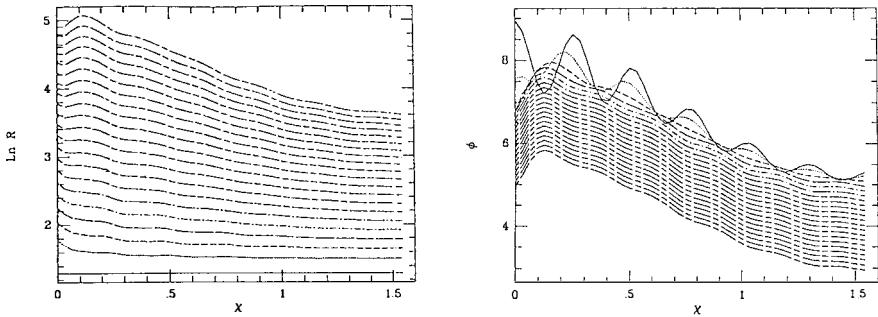
Fig. 3: *The scale factor, R (left) and the scalar field, Φ (right) as a function of the radial coordinate χ for different times. The solid line describes the initial data: large gradients on top of a large $\bar{\Phi}$. We see that the fluctuations in the scalar field decay and inflation starts.*

This analysis suggests that the gradient term decays and does not interfere with the onset of inflation. However, to explore the full effect of initial inhomogeneities we must turn to numerical calculations [11]. A numerical solution of a spherically inhomogeneous Universe with rapid variations of the scalar field is shown in Fig. 3. We see that inflation occurs even in the presence of large gradients, provided that they are superimposed on a large average scalar field ($\bar{\Phi} >$ (a few) $\times M_{pl}$).

The situation becomes more complicated and more interesting when we consider a different inhomogeneous configuration in which at some point, say the origin, $\bar{\Phi} \gg$ (a few) $\times M_{pl}$ while in other regions it is not so large [12]. Figs. 4 and 5 display the evolution of two almost similar Universes that differ in the width, $R\Delta$, of the "effectively homogeneous" region over which Φ is above some critical value. When this region in large the Universe inflates (Fig. 4) but it does not inflate when this region is narrower (Fig. 5). Generally the question whether inflation commences or not depends on the ratio between $R\Delta$ and the horizon size H^{-1}. Fig. 6 displays the expansion at the origin as a function of $R\Delta/H^{-1}$. We see that inflation does not begin unless the scalar field is higher than (a few) $\times M_{pl}$ across several (at least 2) horizons.

Clearly, inflation solves the horizon problem by many orders of magnitude and the initial conditions for inflation are much more general than those required for a Friedmann Universe. But there still remains a problem with initial conditions: *Is it*
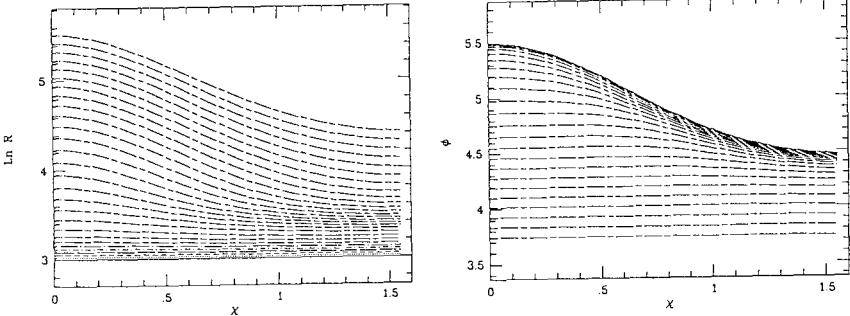
Fig. 4: *The scale factor, R (left) and the scalar field, Φ (right) as a function of the radial coordinate χ for Gaussian with $R\Delta/H^{-1} = 4$. Inflation begins at the origin but not at the exterior region.*
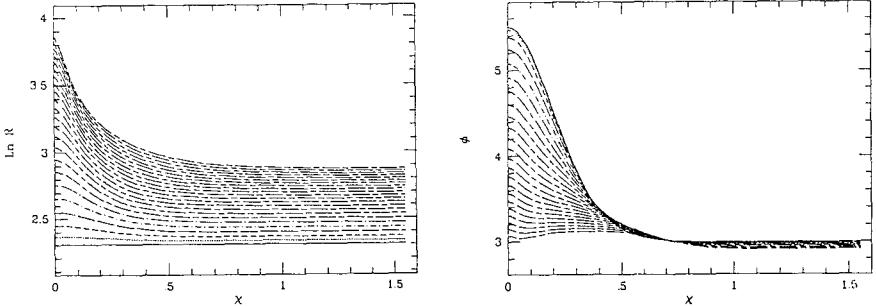
Fig. 5: *The scale factor, R (left) and the scalar field, Φ (right) as a function of the radial coordinate χ for Gaussian with $R\Delta/H^{-1} = .87$ inflation does not appear anywhere.*

reasonable to expect that regions of several horizons over which the average scalar field will have a large value, appropriate for inflation, will exist in the pre-inflationary era?

Suppose that at the end of the quantum era (when $R \approx M_{pl}^{-1}$ and $\rho \approx M_{pl}^4$) the energy of the scalar field is distributed equally between the kinetic, the gradient and the potential terms:

$$\frac{\dot{\Phi}^2}{2} \approx \frac{\delta\Phi^2}{2R^2\Delta^2} \approx \frac{m^2\Phi^2}{2} \approx M_{pl}^4 \quad . \tag{13}$$

The scalar field varies with a typical wavelength $R\Delta \approx M_{pl}^{-1}$ and amplitude $\delta\Phi \approx M_{pl}$. With the assumption of equipartition $\delta\Phi$ is much smaller than the average value of the scalar field, $\bar{\Phi}$. To see this recall that the quantum fluctuation constraint [13]:

$$\delta\Phi/\Phi \approx H/2\pi \ll 10^{-4} \tag{14}$$
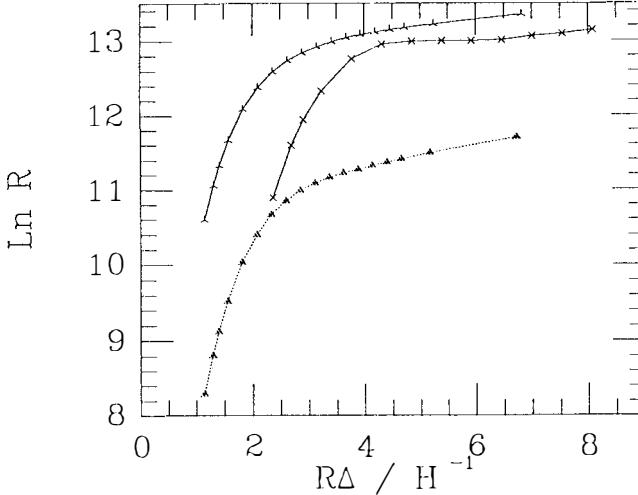
Fig. 6: *The scale factor at the origin at the end of the computations a function of the the proper width of the initial gaussian relative to the horizon size for three different families of initial data and different potentials.*

limits the coupling constant of the scalar field: $m \ll M_{pl}$. The scalar field must have a very large average value,

$$\bar{\Phi} \approx M_{pl}^2/m \gg M_{pl} \tag{15}$$

in order that the potential term will be in equipartition with the kinetic and gradient terms in spite of its small coupling constant. Since $\bar{\Phi} \gg \delta\Phi$, large regions with $\bar{\Phi} \gg$ (a few) $\times M_{pl}$ will exist and inflation can easily start. This can also be viewed differently. The energy density has initially three roughly equal terms, the kinetic term and the gradient terms decay and we are left with a dominant potential term and with inflation.

Our conclusion can be drastically different if we assume that the scalar field Φ emerges from the quantum era in a thermal equilibrium (with $T \approx M_{pl}$) (it has been argued [14] that a weakly coupled scalar field does not have enough time to thermalize during the quantum era, but other workers [15] assert that the scalar field is in a thermal equilibrium during the whole quantum phase). In this case:

$$\frac{\dot{\Phi}^2}{2} \approx \frac{\delta\Phi^2}{2R^2\Delta^2} \approx M_{pl}^4 \quad , \tag{16}$$

and $\delta\Phi \approx M_{pl}$. But the potential energy is much lower than the kinetic energy:

$$\frac{m^2\bar{\Phi}^2}{2} \approx \frac{m^2}{T^2}M_{pl}^4 \quad , \tag{17}$$

and $\bar{\Phi} \approx M_{pl}$ (with $R \approx 1/M_{pl}$. $\bar{\Phi} \approx \delta\Phi$ for a thermal field at $T \gg m$ and we do not expect to find the required large regions with $\bar{\Phi} \gg$ (a few) $\times M_{pl}$. In other words, the

kinetic term and the gradient term are much larger than the potential term. The initial evolution is dominated by them and it is quite likely that the universe will collapse before they decay. In any case by the time that they are comparable to the potential term the scalar field is not large enough to lead to inflation.

We have seen that two different, but plausible, arguments have led us to opposite conclusions. We are faced once more with the question of initial conditions. It seems that we must understand better the conditions at the end of the pre-inflationary era, in order to know whether inflation did take place in our Universe. We can turn to quantum gravity and hope that quantum processes favor configurations in which domains larger than the horizon with $\bar{\Phi} > $ (a few) $\times M_{pl}$ appear at the end of the quantum era. However, today, practically nothing is known about inhomogeneous quantum gravity. Lacking a clear theory the statement - "quantum gravity will provide the necessary initial conditions for inflation" seems to be only an "idea for an idea" which, unfortunately, cannot be pursued much further today.

The uncertainty about the initial conditions turns our attention to an attractive alternative - perhaps there were none? Linde [1] has recently pointed out that *eternal inflation* can take place under relatively simple conditions. Consider an inflating Universe. Classically, Φ decreases, in one e-folding time H^{-1}, by:

$$d\Phi_{classical} \approx \frac{M_{pl}^2}{\Phi} \quad . \tag{18}$$

At the same time Φ undergoes quantum fluctuations of the order

$$d\Phi_{quantum} \approx \frac{m\Phi}{M_{pl}} \quad . \tag{19}$$

Evidently, if Φ is large enough, (more specifically if $\Phi > M_{pl}\sqrt{M_{pl}/m}$), the quantum fluctuations are larger than the classical change. The combined effect is a stochastic random walk with an average decrease in Φ. In some regions Φ will decrease but in others it will increase by $d\Phi_{quantum} - d\Phi_{classical}$. The latter regions will expand faster than the former and will contain a larger volume of the Universe. This process repeats itself again and again every e-folding time. It is stochastic and it is impossible to predict where Φ will increase but the overall effect is clear - some regions of the Universe will inflate for ever. Inflation can be eternal.

Once Φ drops somewhere below $M_{pl}\sqrt{M_{pl}/m}$ the classical motion takes over, Φ decreases and this region eventually exits inflation and emerge as Friedmann Universes like the one in which we live. It seems that eternal inflation provides us with a grand steady state cosmology, in which quantum fluctuations overcome the classical evolution to maintain a stochastic kernel of an inflating Meta-Universe that keeps forever forming domains which exit inflation, one of which is our Universe.

The most remarkable feature of eternal inflation is that it can be eternal on both "ends". It does not halt, but there is no need to turn it on either. One can, out of conceptual inertia, assume that the eternally inflating Universe had a pre-inflationary epoch (with an initial singularity or a quantum era) and immediately be faced with the problem that we have just encountered: how did it start? However, this is not necessary. One can just as well assume that there was no beginning (and that there will be no ultimate end) and that we live in a Universe which is a minuscule part of a steady state eternally inflating Meta-Universe.

This seems like a drastic proposal - but for the time being, and most likely for a very long time in the future, there does not seem to be a single observational clue - or even an idea for one - which will enable us to distinguish between a Universe which began with a bang (and has undergone an inflationary phase later) and one which is a part of an ever inflating Meta-Universe. The fact that there is no present observational distinction between these two options is not necessarily a virtue: it be

nice to be able to test this radical proposal. But this also means that eternal inflation without a beginning cannot be ruled out right away. At least for the time being it should be taken as seriously as the (by now more conventional) initial singularity or initial quantum era proposal. Since the assumption of no initial conditions seems to be the simplest one, Okham's razor will tell us to prefer it and to conclude that we live in a tiny part of a steady state inflating Meta-Universe that has existed and will exist forever.

Acknowledgment

REFERENCES

1. A.D. Linde, Phys. Lett. **175B**, 395 (1986). A.D. Linde, "Inflation and Quantum Cosmology", Acad. Press, Boston, (1990). A.D. Linde, "Particle Physics and Inflationary Cosmology", Gordon and Breach, N.Y., (1990).
2. R.H. Dicke and P.J.E. Peebles, in General Relativity: An Einstein Centenary Survey, eds. S. Hawking and W. Israel. Cambridge: Cambridge University Press.
3. R. Brout, F. Englert and E. Gunzing, in Gravitational Award First Prize Essay, (1978); Gen. Relativ. Gravit. **10**, 1 (1979); A.H. Guth, Phys. Rev. **D23**, 347 (1981).
4. A.D. Linde, Phys. Lett. **108B**, 389 (1982); A. Albrecht and P.J. Steinhardt, Phys. Rev. Lett. **48**, 1220 (1982).
5. A.D. Linde, Phys. Lett. **129B**, 177 (1983).
6. V.A. Belinsky,L.P. Grishchuk,I.M. Khalatnikov and Ya.B. Zel'dovich, Phys. Lett. **155B**, 232 (1985); A.D. Linde, Phys. Lett. **162B**, 281 (1985); T. Piran, and R.M. Williams, Phys. Lett. **163B**, 331 (1985);
7. T. Piran, Phys. Lett. **181B**, 238 (1986);
8. V.A. Belinsky, H. Ishihara, I.M. Khalatnikov, and H. Sato, Prog. Theo. Phys. **79**, 676 (1988).
9. D.S. Goldwirth, Phys. Lett. B in press.
10. H. Feldman and R. Brandenberger, Phys. Lett, **227B**, 359, (1989); J. H. Kung and R. Brandenberger, Phys. Rev **D40**, 2352, (1989); J. H. Kung and R. Brandenberger, Brown University preprint, (1990).
11. D.S. Goldwirth and T. Piran, Phys. Rev. **D40**, 3263 (1989).
12. D.S. Goldwirth and T. Piran, Phys. Rev. Lett. in press (1990).
13. V.F. Mukhanov and G.V. Chibisov, **JETP** Lett. **33**, 523 (1981); S.W. Hawking, Phys. Lett. **115B**, 339 (1982); A.A. Starobinsky, Phys. Lett. **117B**, 175 (1982); A.H. Guth and S.-Y. Pi, Phys. Rev. Lett. **49**, 1110 (1982); J. Bardeen, P.J. Steinhardt and M. Turner, Phys. Rev. **D28**, 679 (1983); R. Brandenberger, Rev. Mod. Phys. **57**, 1 (1985); V.F. Mukhanov, **JEPT** Lett. **41**, 493 (1985); V.F. Mukhanov, L.A. Kofman and D.Yu. Pogosyan, Phys. Lett. **193B**,427 (1987).
14. A.D. Linde, Phys. Lett. **132B**, 317 (1983); A.D. Linde, Phys. Lett. **162B**, 281 (1985); G.F. Mazenko, W.G. Unruh, and R.M. Wald, Phys. Rev. **D31**, 273 (1985).
15. R. Brout, G. Horwitz and D. Weil, Phys. Lett. **B192**, 318 (1987).

THE FRACTAL STRUCTURE
OF THE QUANTUM SPACE-TIME

Laurent Nottale

CNRS.
Département d'Astrophysique Extragalactique et de Cosmologie.
Observatoire de Meudon. F-92195 Meudon Cedex. France

ABSTRACT.

We sum up in this contribution the first results obtained in an attempt at understanding quantum physics in terms of non differential geometrical properties.[1] It is proposed that the dependance of physical laws on spatio-temporal resolutions is the concern of a scale relativity theory, which could be achieved using the concept of a fractal space-time. We recall that the Heisenberg relations may be expressed by a universal fractal dimension 2 of all four coordinates of quantum "trajectories", and that such a point particle path has a finite proper angular momentum (spin) precisely in this case $D=2$. Then we comment on the possibility of a geodesical interpretation of the wave-particle duality. Finally we show that this approach may imply a break down of Newton gravitational law between two masses both smaller than the Planck mass.

1. INTRODUCTION

The present contribution describes results obtained from an analysis of what may appear as inconsistencies and incompleteness in the present state of fundamental physics. We give here a summary of the principles to which we have been led and of some of our main results, which are fully described in Ref. 1.

The first remark is that, following the Galileo/Mach/Einstein analysis of motion relativity, the non absolute character of space and space-time appears as an inescapable conclusion.[2] The geometry of space-time should depend on its material and energetic content. However present quantum physics assumes space-time to be Minkowskian, i.e. absolute, while moreover the fundamental behaviour and properties of quantum objects are known to be radically at variance with classical properties, from which the Minkowskian space-time was yet derived.

The second remark is that Einstein's principle of general relativity ("the laws of physics should apply to systems of reference in any state of motion") is still unachieved. It is now considered, in particular, as not applying to quantum motion. To quote Einstein in this connection, "...I am a fierce supporter, not of differential equations, but of the principle of general relativity, whose heuristic strength is essential to us."[3]

The third remark is that the consequences of one of the radically new behaviour of the quantum world relative to the classical one, i.e. the scale (and/or resolution) dependance of physical laws, may still not have been fully drawn. Though an essential part of the quantum theory through the so-called measurement theory, this scale dependance has still not be included into the laws of physics themselves, in spite of its clearly recognized universality based on the Heisenberg relations.

It is clear that a set of physical measurements takes its complete physical sense, even in the classical domain, only when the measurement resolutions or "errors" have been specified. In quantum physics, the result of a momentum measurement depends explicitely, although in a statistical manner, of the spatial resolution, and the result of an energy measurement depends of the temporal resolution with which the measurement has been performed.

We suggest[1,4] that this fundamental scale dependance of physics is relevant, as motion does, of a relativity theory. Our proposal is to introduce explicitely the resolution in physical laws, either as a new coordinate, or better as a state of scale of the coordinate system, in the same way as velocity and acceleration describe its state of motion. The axes of such a generalized coordinate system can be viewed as endowed with thickness. In such a frame one would require general covariance of physical equations, not only on motion, but also on scale.

The hereabove generalization in the definition of coordinate systems, once assumed universal for a consistent description of physical laws, immediately implies a generalization of the nature of space-time itself. We postulate that the scale

dependance of physics in the quantum domain, and more generally the quantum behaviour itself, take their origin into an intrinsic dependance of space-time geometry on resolution. This implies a generalized metric element where the metrics potentials become explicit functions of resolution:

$$ds^2 = g_{\mu\nu}(x^\lambda, \Delta x^\lambda) \, dx^\mu \, dx^\nu \tag{1}$$

The achievement of the hereabove working hypothesis needs the use of an adequate geometrical mathematical tool. We have suggested that the concept of a continuous and self-avoiding fractal space-time might be such an adequate tool.[1]

2. WHY FRACTALS ?

The suggestion that the quantum space-time possesses fractal structure is supported by a lot of converging arguments. First of all, fractals are characterized by an effective and explicit dependance on resolution. Covering a fractal domain of topological dimension T and fractal dimension $D>T$ by balls of radius Δx yields a T-hypervolume measure which diverges when $\Delta x \to 0$ as:

$$V(x, \Delta x) = \xi(x, \Delta x) . (\lambda/\Delta x)^{D-T} \tag{2}$$

where $\xi(x, \Delta x)$ is a finite fractal function and λ is a fractal / non fractal scale transition.

Second, fractals are also characterized by their non-differentiability, while it had be realized by Einstein[3] (see the hereabove quotation) that giving up differentiability could be the price to be paid in order to apply the principle of general relativity to microphysics, and by Feynman[5] that the quantum "trajectories" of particles can be described as continuous but non differentiable curves.

The third point is that fractals contain infinities, which may be renormalized in a natural way.[6] The correspondence between the renormalization group methods, particularly efficient in the domain of the asymptotic behaviour of quantum field theory, and fractals, has already been pointed out by Callan, as quoted by Mandelbrot[7]. In fact the fractal approach may come as a completion of the renormalization group theory.[8]

Our working hypothesis is that the quantum space-time is a fractal self-avoiding continuum whose geodesics define free particle trajectories. The continuum assumption is a conservative choice allowing to keep a field approach and representation, while the self-avoidance assumption is a necessary condition for such a geodesical interpretation. But because of the non-differentiability of a fractal space, the properties of geodesics will be fundamentally different from those of standard spaces (see Ref.1 and Sec.4 hereafter).

The fact that fractal trajectories run backward relatively to classical coordinates also implies a loss of information in the projection from the intrinsic (fractal) to the classical coordinates, which are defined as the result of a smoothing out of the fractal coordinates with balls larger than the fractal/non fractal transition λ (see hereafter and Refs. 1 and 4 for the quantum interpretation of this transition). We have proposed to describe this behaviour by a "fractal derivative":[1]

$$ds/dx = \Sigma_i \, ds_i/dx = \xi'(x,\Delta x).(\lambda / \Delta x)^{D-T} \tag{3}$$

the sum being performed on all proper time (fractal) intervals found in the classical interval dx . The power $D-T$ is found again because both the topological dimension and the fractal dimension (nearly everywhere[9]) are decreased by 1 in the projection .

3. THE FRACTAL DIMENSION OF QUANTUM TRAJECTORIES.

We now sum up some of the main results obtained or reviewed in Ref. 1. The question of the fractal dimension of a quantum mechanical path in the non relativistic case has been considered by Abbott and Wise[10], Campesino-Romeo et al[11] , Allen[12] and others. The result is a universal value, $D=2$, the transition from non-fractal (classical, $D=T=1$) to fractal (quantum, $T=1$, $D=2$) occuring around the de Broglie length, $\lambda_{dB}=\hbar/p$. Let us recall briefly the argument to show how this result is a direct consequence of the position-momentum Heisenberg relation.

The total length travelled in the average during a set of experiments where the successive positions of a particle are measured with a resolution Δx is given by:

$$L \propto |v| \propto <|p|>, \tag{4}$$

in the non relativistic case. The classical case $<|p|> \approx p_0 >> \Delta p$, i.e. $\Delta x >> \lambda_{dB}$ (since $\Delta x.\Delta p \approx \hbar$) yields as expected a length independant from Δx. On the contrary the quantum case $\Delta p < p_o$ (i.e. $\Delta x < \lambda_{dB}$) yields $<|p|> \approx \Delta p$, so that the length diverges as:

$$L \propto \lambda/\Delta x \tag{5}$$

which corresponds with $T=1$ and from Eq. (1), to $D=2$.

The relativistic case is more difficult to deal with. A superficial analysis may lead to the conclusion that, because of the limitation $v \leq c$, the length will become bounded again for $\Delta x < \approx \lambda_c = \hbar/mc$. However one should also account for the radically new physical behaviour which takes place in the quantum relativistic domain, i.e. virtual particle-antiparticle pair creation-annihilation. I have proposed to reinterpret the virtual pairs as a manifestation of the fact that the fractal trajectory is now allowed to run backward in time for time intervals $\Delta t < \approx \tau_{dB} = \hbar/E_o$.[1] This is based on the Feynman/Stueckelberg/Wheeler interpretation of antiparticles as particles running backward in time. However here we consider electron-positron virtual pairs to be part

of the nature of the electron itself, in agreement with the QED expression for the electron self-energy, which contains all successive Feynman graphs.[13] Indeed consider that for a time interval Δt, an energy fluctuation $\Delta E \approx \hbar/\Delta t$ may give rise to the creation of n e^+e^- pairs, with $\Delta E \approx 2nmc^2$. With the hereabove reinterpretation, the total time elapsed on the fractal trajectory is now given by the sum of the proper times of the $(2n+1)$ particles, i.e.:

$$T = (2n+1)\, t_o = (E+\Delta E)\, t_o/mc^2 = (1+\tau_{dB}/\Delta t)\, t_o \qquad (6)$$

So we get a temporal non-fractal $D=1$ to $D=2$ fractal transition around the de Broglie time τ_{dB}.[1] The argument holds also for the total distance travelled, $L=(2n+1)L_o= (1+c\tau_{dB}/\Delta x)ct_o$, which demonstrates that the spatial fractal dimension remains $D=2$ in the relativistic case. Hence the transition from quantum non relativistic to quantum relativistic may be accounted for by a purely temporal transition. The length increase is now compensated by a time increase, which results into the relativistic bound $v \leq c$. Finally we get a Lorentz covariant scheme with a unique spatio-temporal transition from $D=1$ to $D=2$ around $\lambda^\mu=\hbar/p^\mu$, $\mu=0$ to 3. This result may also be obtained by the inverse argument: Starting from the hypothesis of underlying fractal structures, the Heisenberg relations may indeed be found under the same conditions.[1]

One of the interesting consequences of this interpretation is that it precisely accounts for the Zitterbewegung, this oscillatory motion of the center of mass of an electron resulting from the Dirac equation. This effect is known to be the result of interactions between the negative energy and positive energy solutions of the Dirac equation. Though it indeed disappears if one keeps only the positive energy solutions to describe an electron, this does not yield a satisfactory solution to the problem, since such a positive energy electron would be completely delocalized[14]. Conversely one may set the localization assumed (or measured) for the electron, $\Delta x=c\Delta t=\hbar c/E$, and then derive the relative rate of positive and negative energy solutions. One gets:

$$P_-/P_+ \approx [pc/(E+mc^2)]^2 = (E-mc^2)/(E+mc^2) \qquad (7)$$

Now in the fractal model, for each classical time interval we have $2n+1$ segments, $n+1$ running forward and n running backward, so that with $E=(2n+1)mc^2$, one gets $P_-/P_+ =n/(n+1)=(E-mc^2)/(E+mc^2)$, i.e. exactly the QED result. This relation of the Zitterbewegung to the fractal approach will be developed in a forthcoming paper.[8]

4. GEODESICAL INTERPRETATION OF THE WAVE-PARTICLE DUALITY.

Let us define the de Broglie length and time (λ, τ) as geometrical structures of the fractal trajectory of a "particle": They are identified to the fractal/non fractal

transition. Then the various classical quantities may be expressed in term of these two geometrical quantities:

$$E = \hbar/\tau \; ; \; p = \hbar/\lambda \; ; \; v = c^2\tau/\lambda \; ; \; m = \hbar c^{-1}[c^{-2}\tau^{-2} - \lambda^{-2}]^{1/2} \tag{8}$$

This means that we do not have to endow the point particle with mass, energy, momentum or velocity, but instead that these properties may be reduced to geometrical structures of the resolution space. Hence the energy-momentum tensor writes in terms of the de Broglie periods and of the Planck length λ_P (with $X_i(S)$ being four fractal functions defining the trajectory of particle i):

$$\frac{G}{c^2} T^{\mu\nu} = \Sigma_i \int c\tau_i \; \frac{\lambda_P}{\lambda_i{}^\mu} . \frac{\lambda_P}{\lambda_i{}^\nu} \; \delta^4[x - X_i(S)].(dt/d\sigma).d\sigma \tag{9}$$

We have also demonstrated that the quantum spin itself may also be obtained as a proper angular momentum of fractal trajectories having precisely fractal dimension 2.[1] While for $D<2$ and $D>2$ the proper angular momenta of fractal curves are respectively 0 and infinite, it becomes finite for the strict value $D=2$ and then owns all the properties of an *internal* quantum number.

Let us now comment on the wave-particle duality. A fractal space-time is characterized by an infinity of obstacles at all scales and by returns and eddies also at all scales and for all 4 coordinates. As a consequence one expects that an infinity of geodesics will connect any two space-time events, so that the properties of physical objects on a fractal space-time will result in a mixing of individual properties (one particular geodesic) and collective properties (those of families of geodesics), in agreement with the wave-particle duality. While one may admit, without any logical inconsistency, that a particle follows one particular geodesic (as indicated by individual *measurements* of well defined particles), one must admit at the same time that all geodesics are equiprobable, so that any *prediction* can be only of statistical nature, since applying to a family of geodesics. So in such a frame we get the possibility, at least in principle, to have both a space-time which would be strictly determined by its material and energetic content, and non deterministic particle trajectories. But here the statistical behaviour would not be the primeval stone on which physical laws are based, but instead a consequence of the (fractal) nature of space-time.

One result supporting this interpretation is that indeed beams of geodesics may have a wave nature described by a Schrödinger-like equation in a Riemannian space-time, even at the geometric optics approximation.[1,15] A light beam in general relativity is described at the geometric optics approximation by a congruence of null geodesics, the equations of which have been written by Sachs.[16] The cross sectional area A of the light beam may be subjected on its way to three infinitesimal

deformations, expansion $\theta.d\omega = d\sqrt{A}/\sqrt{A}$, rotation $\Omega.d\omega = dW$ and shear. Considering only the shearless case, when setting $\psi = \sqrt{A}.e^{iW}$ the beam propagation equation writes, in term of an affine parameter ω and of Ricci tensor R_{ij} and wave vector k^i:[1]

$$d^2\psi / d\omega^2 - \frac{1}{2}R_{ij}.k^i.k^i \; \psi = 0 \qquad (10)$$

which has exactly the form of the one-dimensional Schrödinger equation. So a family of geodesics, even at the geometric optics approximation, possesses the equivalent of a quantum phase, interpreted as the beam rotation and of a probability of presence, identified to the beam cross sectional area.

5. POSSIBLE IMPLICATIONS.

Let us consider a simple model of metrics accounting for the constraints which have been hereabove obtained:

$$ds^2 = \xi^2(t,\Delta t) \; (1-\tau/\Delta t)^2 \; c^2 \; dt^2 - dl^2 \qquad (11)$$

in which ξ is some finite fractal function (with $\xi = 1$ for $\Delta t > \tau$), $\tau = \hbar/mc^2$ is the de Broglie time of a system of inertial mass m and dl^2 is a fractal spatial element which will not be more detailed here. This form of g_{oo} has been chosen for the following reasons: It yields fractal dimension 2 for $\Delta t \ll \tau$, the Minkowski limit $g_{oo} = 1$ for $\Delta t \to \infty$, and a singularity $g_{oo} = 0$ for $\Delta t = \tau$, thus accounting for creation/annihilation of particles (in analogy to the cosmological primeval singularity). Consider the domain $\Delta t \gg \tau$, g_{oo} may be expanded as $g_{oo} = 1 - 2\tau/\Delta t$. Let us now identify $\Delta t \equiv r/c$ to the lifetime of an exchanged virtual boson. One gets:

$$g_{oo} = 1 - 2\hbar/mcr \qquad (12)$$

This result holds for one particle of mass m, but also for a complex system of total mass Σm, thanks to the universality of the de Broglie wave nature of any physical system . However one may remark that the wave properties of a system keep a physical sense only if one may, at least in a "gedanken experiment", measure them (i.e. by a diffraction or interference experiment). But when the total mass m becomes larger than the Planck mass $m_P = (\hbar c/G)^{1/2} \approx 2 \; 10^{-5}$ g, its Compton length becomes smaller than its Schwarzschild radius $r_s \approx Gm/c^2$, thus becoming unmeasurable, not only for technological limitation, but mainly for a profound physical limitation, since it enters into a black hole horizon.[1] Conversely for $m < m_P$, the Schwarzschild radius becomes smaller than the Compton wavelength, inside which the concept of position itself loses its physical meaning,[13] while a black hole state can be reached only provided the whole mass m is confined into its Schwarzschild radius. So we suggest that both the hereabove potential $g_{oo} = 1 - 2\hbar/mcr$ and the Schwarzschild potential

20

$g_{oo}=1-2G/c^2r$ become unphysical, respectively for $m>m_P$ and $m<m_P$. The remarkable result here is that *they are precisely equal for m=m_P*.

So our proposal is that there exists some fundamental microscopic/macroscopic transition around the Planck mass, which would hold at macroscopic scale and would connect the quantum behaviour to the gravitational (general relativistic) one. This is achieved by the following phenomenological potential:

$$g_{oo} = 1-2(\frac{m}{m_P}+\frac{m_P}{m})\frac{\lambda_P}{r} \qquad (13)$$

As an observable consequence, the Newton law should break down between two dust particles both having mass smaller than the Planck mass.[1] This could be checked in a space experiment : For example two 2.10^{-5} g dust grains of density d=18 have a radius of 0.064 mm. Their expected Newtonian free fall time for an initial relative distance 0.24 mm would be ≈10 mn. We expect the falling time to become smaller than the Newtonian time for $m<\approx m_P$. Such an experiment would be at the limit of the present possibilities, needing a high control of vessel gravity gradients and of electric charges, since the electric force would be equal to the m_Pxm_P gravitational one for only $\alpha^{-1}\approx12$x12 elementary electric charges.

The hereabove calculation is a very rough one including only temporal terms, in which the mass considered was assumed to remain a point mass. We intend to attempt its spatio-temporal generalization to the extended case, which should lead us to the generalization of the Planck mass transition to a critical density transition.[8] Then astrophysical and cosmological implications, concerning e.g. the formation of structures, are expected to be derived in such a way that the hereabove suggestion of a transition between a quantum and a gravitationnal regime should be falsifiable.

REFERENCES

1. Nottale, L., 1989, *Int. J. Mod. Phys.* **A4**, 5047
2. Einstein, A., 1916, *Annalen der Physik* **49**,769
3. Einstein, A., Letter to Pauli, 1948, in *Albert Einstein, Quanta*, Seuil/CNRS, p.249
4. Nottale, L., 1988, *C.R. Acad. Sci. Paris* **306**, 341
5. Feynman, R. & Hibbs, A., *Quantum mechanics and path integrals* (McGraw-Hill, NY, 1965)
6. Nottale, L., & Schneider, J., *J. Math. Phys.* **25**, 1296
7. Mandelbrot , B., *The fractal geometry of nature* (Freeman, San Francisco, 1982) p.331
8. Nottale, L., in preparation
9. Mandelbrot, B., *The fractal geometry of nature* (Freeman, San Francisco, 1982) p.365
10. Abbott, L.F. & Wise, M.B., 1981, *Am. J. Phys.* **49**, 37
11. Campesino-Romeo, E., D'Olivo, J.C. & Socolovsky, M.,1982, *Phys. Lett.* **89A**, 321
12. Allen, A.D., 1983, *Speculations in Science and Technology* **6**,165
13. Landau, L. & Lifshitz, E., *Relativistic Quantum Theory*, (Mir, Moscow, 1972)
14. Bjorken, J.D. & Drell, S.D., *Relativistic Quantum Mechanics*, (McGraw-Hill, NY, 1964)
15. Nottale, L., 1988, *Ann. Phys. Fr.* **13**, 223
16. Sachs, P.K., 1961, *Proc. Roy. Soc. London* **A264**, 309

STATISTICAL PROPERTIES OF CURVATURE PERTURBATIONS
GENERATED DURING INFLATION

Sabino Matarrese

Dipartimento di Fisica *G.Galilei*, Università di Padova
via Marzolo 8, I–35131 Padova, Italy

ABSTRACT

The stochastic dynamics of the scalar field responsible for inflation is considered in connection with the statistical properties of classical curvature perturbations which are generated by quantum fluctuations in that field. The combined effect of non–linearities in the scalar field and of perturbations in the metric makes curvature perturbations on large scales strongly non–Gaussian. The controversial issue of whether perturbations in our observable patch of the inflated universe are also non–Gaussian is discussed in terms of conditional probabilities.

INTRODUCTION

Density fluctuations generated during inflation (see, e.g., the review by A. Linde in these Proceedings) are usually considered to be Gaussian, as a general consequence of the required flatness of the potential for the *inflaton*, the scalar field which drives the accelerated universe expansion. Recently, however, it has been shown that both isocurvature [1)−4)] and curvature perturbations [5)−9)] can be characterized by non−Gaussian statistics. Due to the back−reaction of field fluctuations on the background geometry, phase correlations appear during the stochastic evolution of horizon−size inflaton modes, thus providing non−Gaussian initial conditions for the linear evolution of adiabatic perturbations. In fact, it has been shown [6)] that, for a wide class of potentials leading either to Linde's chaotic inflation [10)], •r to power−law inflation [11),12)], fluctuations in the scalar field are non−Gaussian distributed around the classical trajectory: at the end of inflation the distribution for the gravitational potential fluctuations can be highly non−Gaussian [9)]. This is an important conclusion for theoretical cosmology since it opens the possibility of unexplored models for the formation of cosmic structures which, abandoning the random−phase paradigm, preserve the simplicity of the gravitational instability picture. Generally speaking one is faced with a new class of models which imply more structure on large scales than the standard cold dark matter model. Models of this type are presently under investigation in connection with their clustering properties on large scales [13)].

The mechanism for the generation of non−Gaussian adiabatic perturbations, which is discussed here is characterized by the absence of intrinsic lengths on cosmologically relevant scales, it therefore implies a *scale−invariant* fluctuation field (see, e.g., [14),15)]). Such a scale−invariance property is properly expressed by a simple scaling (up to negligible logarithmic corrections) of the peculiar gravitational potential $\Phi(\mathbf{x})$ in Fourier space at every time during matter dominance

$$\langle \Phi(\mu\mathbf{k}_1)\dots\Phi(\mu\mathbf{k}_N)\rangle d^3(\mu\mathbf{k}_1)\dots d^3(\mu\mathbf{k}_N) \approx \mu^{N(n_p-1)/2}\langle\Phi(\mathbf{k}_1)\dots\Phi(\mathbf{k}_N)\rangle d^3 k_1 \dots d^3 k_N \quad (1)$$

where n_p is the primordial spectral index. For $n_p = 1$, Eq.(1) represents a generalization of the Zel'dovich criterion of scale−invariance to non−Gaussian fluctuations. This immediately translates into a statement on the time evolution of the N−point correlation functions $\xi^{(N)}(\mathbf{x}_1,\dots,\mathbf{x}_N;t)$ for mass fluctuations, consistently defined in terms of the Zel'dovich approximation [16)], up to the time of first shell−crossing

$$\xi^{(N)}(\mathbf{x}_1,\dots,\mathbf{x}_N;t_0) \approx \xi^{(N)}(\mu\mathbf{x}_1,\dots,\mu\mathbf{x}_N;t) \quad (2)$$

provided that $\mu = [b(t)/b(t_0)]^{2/(n_p+3)}$, with b the growing mode of linear perturbations, proportional to $t^{2/3}$ in the matter dominated era and in a flat universe. This last property, however, only applies over suitably large scales where the curvature of the primordial spectrum introduced during the linear evolution of perturbations inside the horizon is unimportant. Although the potential Φ is simply related to the linear density fluctuation $\delta\varrho$ through the

Poisson equation

$$\triangle^{(3)}\Phi(\mathbf{x}, t) = -4\pi G a^2(t)\delta\varrho(\mathbf{x}, t), \tag{3}$$

with $a(t)$ the scale–factor, we prefer to define the mass density through the Zel'dovich approximation; this allows to extend the treatment to the mildly non–linear evolution and to take into account the constraint $\varrho(\mathbf{x}) \geq 0$. Simple toy–model examples of scale–invariant statistics are: the model proposed by [17], where the density fluctuation field is the convolution of two independent scale–free Gaussian processes; a model where the density field is the square of a Gaussian random process [as recently found by Bardeen (unpublished) in the analysis of a two–scalar field model for inflation]. As we shall see in the following, the non–Gaussian scale–invariant density fluctuations obtained from inflation are all of multiplicative type (see also [7]); this is an important property since a number of interesting phenomena are likely to occur in this case. Among these there is an interesting phenomenon called *intermittency* (see, e.g., [18]), which is clearly exhibited for instance by log–normal random fields; roughly speaking, intermittency consists in the occurrence, in realizations of the random field, of sporadic high spots where most of the intensity is stored, separated by large regions of reduced intensity.

A quite different mechanism for producing non–Gaussian and non–scale–invariant perturbations during inflation rests on the use of multiple (interacting) scalar fields during inflation as considered in refs. [4],[7]; the stochastic method has been recently extended to the multiple scalar field case (see, e.g., [19],[20]). In this case one can easily obtain non–Gaussian and/or non–scale–invariant perturbations both of adiabatic and isocurvature type (see, e.g., the contribution by S. Mollerach to these Proceedings).

STOCHASTIC INFLATION

It has been shown by many authors that the dynamics of the inflaton on scales larger than the comoving Hubble radius $r_H(t) \approx 1/\dot{a}(t)$ is accurately described by a stochastic approach (see, e.g., [21],[22]). One defines a *coarse–grained* variable $\varphi_\mathbf{x}(t)$ which is the average of the quantum scalar field over a volume of size $\sim r_H^3(t)$. In the spirit of the chaotic inflation scenario initial conditions are introduced by assuming that there are initial domains of size $\sim r_H^3(t_\star)$ characterized by a nearly homogeneous value $\varphi_\star$ for the scalar field. Provided that $\varphi_\star$ is large enough, a few Hubble times suffice to depress both the kinetic energy and the spatial gradients compared to the potential energy $V(\varphi)$. The resulting coarse–grained dynamics is *friction* dominated and can be described by a Langevin–type equation

$$\dot{\varphi}_\mathbf{x} = -\frac{1}{3H(\varphi_\mathbf{x})}\frac{\partial V(\varphi_\mathbf{x})}{\partial\varphi_\mathbf{x}} + \frac{H^{3/2}(\varphi_\mathbf{x})}{2\pi}\eta_\mathbf{x}(t), \tag{4}$$

where $H(\varphi_\mathbf{x}) \approx \sqrt{V(\varphi_\mathbf{x})/3\sigma^2}$ and $\sigma \equiv 1/\sqrt{8\pi G}$. In this approach $\mathbf{x}$ labels the coarse–grained variable in different cells. In writing Eq.(4) one implicitly assumes a perturbed Friedmann line–element in an appropriate synchronous gauge (see, e.g., [23])

$$ds^2 \approx dt^2 - a^2(\mathbf{x}, t)dl^2, \tag{5}$$

with a local scale–factor

$$a(\mathbf{x}, t) \approx a(\mathbf{x}, t_\star) \exp \int_{t_\star}^{t} H(\varphi_\mathbf{x}) dt'. \tag{6}$$

Non–diagonal scalar perturbations of the metric, which are also initially present in this gauge, are quickly depressed, on scales larger than the horizon, by the inflationary expansion. The first term on the r.h.s. of Eq.(4) plays the role of a classical convective force, while the second one represents the diffusion process induced by fine–grained quantum fluctuations. The dependence of H on φ makes the stochastic process a multiplicative one. The noise $\eta_\mathbf{x}(t)$, which has zero mean, is accurately approximated by a stationary Gaussian process with auto–correlation function

$$\langle \eta_\mathbf{x}(t) \eta_{\mathbf{x}'}(t') \rangle = \hbar \, j_0 \big(|\mathbf{x} - \mathbf{x}'|/r_H(t)\big) \, \delta(t - t'), \tag{7}$$

where j_0 is the zero order spherical Bessel function. Because of the *white–noise* character of η, with respect to its time dependence, $\varphi(t)$ is *Markovian*. The quantity $\mathcal{P}(\varphi, t) \, d\varphi = \langle \delta(\varphi - \varphi[\eta_\mathbf{x}(t)]) \rangle_\eta \, d\varphi$, yielding the (*one–particle*) probability that, in a randomly chosen point $\mathbf{x}$, $\varphi_\mathbf{x}$ takes a value in the interval $\varphi, \varphi + d\,\varphi$, evolves according to a Fokker–Planck equation (in the Einstein–Smoluchowski limit)

$$\frac{\partial \mathcal{P}}{\partial t} = \frac{\partial}{\partial \varphi} \left[\left(\frac{2\sigma}{3^{1/2}} \frac{\partial V^{1/2}}{\partial \varphi} \right) \mathcal{P} + \frac{\hbar}{4\pi^2} \frac{V^{3/4}}{3^{3/2} \sigma^3} \frac{\partial}{\partial \varphi} \left(V^{3/4} \mathcal{P} \right) \right]. \tag{8}$$

Once the inflaton potential $V(\varphi)$ has been specified, one looks for a time–dependent solution with the initial condition $\mathcal{P}(\varphi, t_\star) = \delta(\varphi - \varphi_\star)$ corresponding to a homogeneous configuration $\varphi_\star$ with potential energy $V_\star \approx 3\sigma^2 H_\star^2$ (one needs $H_\star/\sigma \lesssim 8\pi/\sqrt{3}$, in order not to exceed the Planck energy).

It should be recalled that the solution of Eq.(8) with a delta–like initial condition is actually a conditional probability (also called *transition probability* in this context): it gives the probability that our stochastic process takes the value φ at time t (in a randomly selected position) given that the result of a measurement was $\varphi_\star$ at time $t_\star$ (Smoluchowski called it *probability after effects*). Because of the Markovian character of the process the sharp condition at $t_\star$ cancels any memory of the evolution preceding $t_\star$; also, for suitably short time intervals after $t_\star$ the form of the probability is strongly dominated by the constraint, while for times long enough the initial distribution is essentially forgotten.

The Langevin equation (4), together with the η correlation function Eq.(7), actually contains much more information than the (one–particle) probability $\mathcal{P}$, for it takes into account the space–correlation properties of the distribution, that is, it allows to obtain the whole probability density functional. The probability density functional evolves according to a suitable *functional Fokker–Planck equation* [such a complete treatment has been sketched, for instance, by Rey [24], although he completely disregarded the role of metric perturbations]. Another effect which should be taken into account when dealing with the space–distribution of the coarse–grained field is connected with the different weights to be assigned to different

cells due to fluctuations in the coarse–graining volume. This problem is dealt with by many authors by multiplying the probability by the volume factor $a^3(\mathbf{x}, t)$, obtained from Eq.(6). This correction is at the origin of the so–called *eternally existing self–reproducing inflationary universe* (see, e.g., [23),25)]); for our purposes neglecting such a correction is a minor approximation since this is expected to mainly affect the global properties of the universe and the evolution of the probability distribution at early times, when the local fluctuations of the Hubble constant can be very large.

It has been shown [6)] that the dynamics described by Eq.(8) [or Eq.(4)] presents universal properties for a large class of models leading either to chaotic or to power–law inflation. These can be summarized as follows. At early times the diffusion term in Eq.(8) causes the initial delta function to spread around its maximum which starts moving, due to the convective term. Convection, at this stage, can be approximated by a constant positive force (namely $-V'/3H$ evaluated at $\varphi_\star$), later on, however, force gradients start to be felt: *Brownian particles* which are closer to the minimum of the potential suffer smaller attraction compared to more distant ones; at the same time the diffusion coefficient becomes smaller and smaller as the minimum is approached and can be safely neglected. This causes a shrinking of the distribution at late times characterizing the so called *scaling regime* (see, e.g., [26)]): at the *scaling time* t_{sc} the system undergoes a transition from a diffusion dominated regime (due to quantum fluctuations) to a convection dominated regime (due to classical non–linearities), or from a *disordered phase* to a *macroscopically ordered phase*. In this picture inflation is described as the non–equilibrium decay of the system from the unstable state $\varphi_\star$ to the minimum of the potential, with the whole coarse–grained field distribution undergoing slow-rolling down. The peaking of the probability at the onset of the scaling regime was noticed by a number of authors [24),27),28)]. During the scaling regime the distribution is strongly peaked around the classical homogeneous configuration $\varphi_{cl}(t)$ [(the solution of Eq.(4) when η is set to zero, or when $\hbar \to 0$], with fluctuations giving rise to classical curvature perturbations. In this *scaling limit* the solution tends to a self–similar time–independent function $\mathcal{W}$ of a single *scaling variable* ξ: $\mathcal{P}(\varphi, t)\, d\varphi \sim \mathcal{W}(\xi)\, d\xi$. Such a scaling variable can be chosen so that it is, with very good accuracy (or even exactly), Gaussian with zero mean and two–point function

$$\langle \xi_{\mathbf{x}}(t)\xi_{\mathbf{x}'}(t') \rangle = \frac{1}{2\pi^2} \int_{1/r_H(t_\star)}^{1/r_H(t_{min})} dk\, k^2 (P_0/k^3) j_0\big(k|\mathbf{x} - \mathbf{x}'|\big), \tag{9}$$

where $t_{min} = min(t, t')$. Only wavelengths that left the horizon during the interval $t_\star \div t_{min}$ contribute to the integral in Eq.(9). The ξ power–spectrum has a *flicker–noise* form $P(k) = P_0 k^{-3}$. The amplitude P_0 (which may possibly contain a residual logarithmic k–dependence) as well as the non–linear transformation between φ and ξ depend upon the precise form of the inflaton potential.

INFLATIONARY MODELS

We shall restrict our discussion to two simple models: chaotic inflation driven by a quartic inflaton potential [10)] and power–law inflation based on an exponential potential, as

first proposed in ref. [12]. Chaotic inflation can be driven by a quartic potential $V(\varphi) = (\lambda/4)\varphi^4$. The classical solution of the φ dynamics in the slow–rolling down phase is described by $\varphi_{cl}(t) = \varphi_\star \exp(-2\sqrt{\lambda/3}\,\sigma\Delta t)$, where $\Delta t = t - t_\star$. If the expansion is dominated by the classical homogeneous mode, the scale–factor is given by

$$a(t) = a_\star \exp\left\{ (H_\star/4\sigma)(3/\lambda)^{1/2}\left[1 - \exp(-4\sqrt{\lambda/3}\,\sigma\Delta t)\right]\right\}. \tag{10}$$

Inflation is expected to end when the kinetic energy equals the potential energy, i.e. when $\varphi = \pm 2\sqrt{2}\sigma$. The exponential potential $V(\varphi) = V_\star \exp[-\lambda(\varphi - \varphi_\star)/\sigma]$, leads to power–law inflation

$$a(t) = a_\star\left[1 + \lambda^2 H_\star\Delta t/2\right]^{2/\lambda^2}, \tag{11}$$

for any $\lambda < \sqrt{2}$. Unless the potential is suitably modified to allow for reheating to occur, inflation lasts forever, the ratio of kinetic and potential energy being constant along the classical trajectory $\varphi_{cl}(t) = \varphi_\star + (2\sigma/\lambda)\ln(1 + \lambda^2 H_\star\Delta t/2)$.

For the quartic potential one can solve exactly the Langevin equation and find, for $\delta\varphi_{\mathbf{x}}(t) = \varphi_{\mathbf{x}}(t) - \varphi_{cl}(t)$ [$\varphi_{cl}(t)$ generally differs from $\langle\varphi(t)\rangle$ by a small, time–dependent, space–homogeneous quantity which does not affect observable quantities]

$$\delta\varphi_{\mathbf{x}}(t) = \pm\, \sigma\left(\frac{12H_\star^2}{\lambda^2\sigma^2}\right)^{1/4}\exp[-2\sqrt{\lambda/3}\,\sigma\Delta t]\left[|1 + \xi_{\mathbf{x}}(t)|^{-1/2} - 1\right], \tag{12}$$

where the solution with the plus (minus) sign must be taken if $\varphi_\star$ is positive (negative). In this case $P_0 \approx \hbar(\lambda/3)^{1/2}(H_\star/\sigma)[1 - 4(\lambda/3)^{1/2}(\sigma/H_\star)\ln(k/a_\star H_\star)]$. For the exponential potential appropriate use of the *scaling approximation* yields, for large times

$$\delta\varphi_{\mathbf{x}}(t) = \frac{4\sigma}{\lambda^3 H_\star t}\left[|1 + \xi_{\mathbf{x}}(t)|^{2/3} - 1\right] \tag{13}$$

and $P_0 = (9\hbar/16)[\lambda^2/(2 - \lambda^2)](H_\star/\sigma)^2$. The random–phase approximation in this context would amount to expand the r.h.s. of Eqs.(12) and (13) to first order in ξ; however ξ can have a large r.m.s. value, depending on the value of $H_\star$, so that such an approximation would fail in the general case. In the exponential potential model, for instance,

$$\xi_{rms}(t) = \left(\frac{\hbar}{2\pi^2}\right)^{1/2}\frac{3H_\star}{4\sigma}\ln^{1/2}(1 + \lambda^2 H_\star\Delta t/2), \tag{14}$$

which, for times after the onset of the scaling regime $\Delta t_{sc} \approx 2(\sqrt{e}-1)\lambda^{-2}H_\star^{-1}$, can be larger than unity. This fact is at the origin of the non–Gaussian behaviour of $\delta\varphi$ discussed by Matarrese, Ortolan and Lucchin [6]. For the quartic potential model, for times much larger than $\Delta t_{sc} \approx (3/\lambda)^{1/2}(1/8\sigma)\ln 3$, due the the large value of $\xi_{rms}(t)$, one can approximately write $\varphi_{\mathbf{x}}(t) \propto |\xi_{\mathbf{x}}(t)|^{-1/2}$. Positive moments of φ, being related to negative moments of a semi–Gaussian distribution (i.e., the distribution for $|\xi|$), are infinite (as for the Cauchy distribution), except for the first one, $\langle\varphi\rangle$. One finds that the probability of crossing a level ν

times a suitable *effective dispersion* $\langle \delta\varphi^2_{eff} \rangle^{1/2}$ is far from the Gaussian expectation already for small ν. This result holds, in the scaling regime, for suitably large values of $H_\star/\sigma$, independently of the value of λ and of time [in disagreement with Hodges [29)], who incorrectly used a random–phase approximation to estimate these crossing probabilities], as follows from the scale–invariance property

$$\mathcal{P}(\varphi, t) \approx \mu \mathcal{P}(\mu\varphi, t - \sqrt{3/4\lambda\sigma^2} \ln \mu). \tag{15}$$

For the exponential potential one finds at large times $\varphi_{\mathbf{x}}(t) \propto |\xi_{\mathbf{x}}(t)|^{2/3}$ and

$$\langle \varphi^N \rangle \sim \pi^{-1/2} \Gamma\left(\frac{1}{2} + \frac{N}{3}\right) \left(\frac{2^{7/3}\sigma}{\lambda^3 H_\star t}\right)^N \xi_{rms}^{2N/3}(t), \tag{16}$$

with ξ_{rms} given in Eq.(14). In the scaling regime the N–th connected moments normalized to the power $N/2$ of the variance relax to constant non–zero values: clear evidence for non–Gaussian scale–invariant behaviour of the statistics. These dimensionless ratios do not depend neither on the value of λ, nor on time (scale), nor on the initial condition through $H_\star$. The role of $H_\star$, however, is more subtle: it must be large enough so that the system has well entered the scaling regime when the cosmologically interesting scales leave the inflationary horizon. It is clear that a low value of $H_\star$ (just enough to solve the horizon problem), as that quoted by Kofman *et al.* [7)], keeps the system in the diffusive regime, where the fluctuations are practically Gaussian, and implies that our universe is exactly homogeneous on scales larger than the Hubble radius. This property is much stronger than what required by the isotropy of the cosmic microwave background, which only demands (for reasonable fluctuation spectra) that the r.m.s. density fluctuation on the scale of the present horizon is less than about 10^{-5}. Fluctuations on super–horizon scales only determine the local value of *average quantities.*

CURVATURE FLUCTUATIONS

For scale–free inflaton potentials such as those presented in the previous section these results can be expressed in an a simple form as follows: the scaling approximation amounts to replacing the Langevin equation (4) by the following effective equation

$$\dot{\varphi}_{\mathbf{x}}(t) = F_{cl}(t) \, \varphi_{\mathbf{x}}(t) + c \, \varphi^\beta_{\mathbf{x}}(t) \, \eta_{\mathbf{x}}(t), \tag{17}$$

where $F_{cl}(t)$ is some function of time, c is a constant and the power β depends upon the inflaton potential. This equation is exactly solved by

$$\varphi_{\mathbf{x}}(t) = \varphi_{cl}(t)|1 + \xi_{\mathbf{x}}(t)|^{1/(1-\beta)}, \tag{18}$$

where $\varphi_{cl}(t)$ is determined by $F_{cl}(t)$ through

$$\varphi_{cl}(t) = \varphi_\star \exp \int_{t_\star}^t dt' F_{cl}(t') \tag{19}$$

[except for the exponential potential case, where the situation is more involved as shown by Eq.(13), Eq.(19) is actually the definition of $F_{cl}(t)$] and

$$\xi_{\mathbf{x}}(t) = (1 - \beta) \, c \int_{t_\star}^{t} dt' \varphi_{cl}^{\beta-1}(t') \, \eta_{\mathbf{x}}(t'). \tag{20}$$

In this simplified treatment the quartic potential, for which this approach is exact, corresponds to $\beta = 3$; more in general, inflaton potentials of the type $V \propto \varphi^{2n}$ are described by equation (17) with $\beta = 3n/2$ and the exponential potential, at late times, corresponds to $\beta = -1/2$. Note that for $\beta = 1$ the solution of this equation yields a log–normal process, while if β goes to zero in the equation (in the absence of boundary conditions) one gets a Gaussian process with $\delta\varphi_{\mathbf{x}}(t)/\varphi_{cl}(t) = \xi_{\mathbf{x}}(t)$. Also the model considered by Bardeen which is derived for two interacting scalar fields is formally described by this equation for $\beta = 1/2$. Therefore Eq.(17) describes a whole set of models, parametrised by β, which goes from the Gaussian case $\beta = 0$ to the extreme case of multiplicative stochastic process $\beta = 1$, leading to the intermittency phenomenon: the log–normal process.

The coarse–grained field fluctuation $\delta\varphi_{\mathbf{x}}(t)$ contains all wavelengths larger than the horizon size at the time t. In order to obtain the peculiar gravitational potential $\Phi(\mathbf{x})$ it is necessary to Fourier transform $\delta\varphi$ at each time t keeping only that mode $k \approx 1/r_H(t)$ that crosses the Hubble radius then. We can obtain the Fourier modes of the gravitational potential fluctuation field which reentered the horizon during the matter dominated era, by using their approximate constancy (the different behaviour of fluctuations which entered the horizon during radiation dominance is properly accounted for by the transfer function). One has

$$\Phi(\mathbf{k}) \approx -\frac{3T(k)}{5} \frac{H_{cl}(t_1)}{\dot{\varphi}_{cl}(t_1)} \, \delta\varphi(\mathbf{k}, t_1), \tag{21}$$

where t_1 is the time when the wave–number k left the horizon during inflation and $T(k)$ is the transfer function appropriate for the type of scenario one is considering. One will therefore find

$$\Phi(\mathbf{k}) \approx \frac{3T(k)}{10} \left(\frac{3H_\star^2}{\lambda\sigma^2}\right)^{1/2} \exp(-4\sqrt{\lambda/3}\ \sigma\ t_1) \int d^3x e^{-i\mathbf{k}\cdot\mathbf{x}} \big[|1 + \xi_{\mathbf{x}}(t_1)|^{-1/2} - 1\big], \tag{22}$$

for the quartic potential and

$$\Phi(\mathbf{k}) \approx -\frac{12T(k)}{5\lambda^4 H_\star t_1} \int d^3x e^{-i\mathbf{k}\cdot\mathbf{x}} \big[|1 + \xi_{\mathbf{x}}(t_1)|^{2/3} - 1\big], \tag{23}$$

for the exponential potential, at times much larger than the scaling time. It is clear that the non–linear transformation from the Gaussian variable ξ to $\delta\varphi$ implies that all Fourier modes of ξ contribute to $\Phi(\mathbf{k})$. As we have shown before, in the scaling regime one can approximately write $\big[|1 + \xi_{\mathbf{x}}(t)|^\alpha - 1\big] \sim |\xi_{\mathbf{x}}(t)|^\alpha$ in the r.h.s. of Eqs.(22) and (23) where $\alpha = 1/(1 - \beta)$ is respectively $-1/2$ and $2/3$. Note that this is the opposite of the random–phase approximation, valid in the initial diffusion dominated regime, where one writes $\big[|1 + \xi_{\mathbf{x}}(t)|^\alpha - 1\big] \sim \alpha\ \xi_{\mathbf{x}}(t)$.

Of course, if one had linearized from the beginning the evolution equation for $\delta\varphi$, the coarse-grained fluctuations would come out proportional to ξ and then Gaussian (non–Gaussian fluctuations can be obtained in the linear approximation by interacting multiple scalar fields, as in ref. [4]: the linear approximation for $\delta\varphi$, therefore, does not adequately follow the dynamics during the convective regime. The non–linear transformation $\xi \rightarrow \Phi(\mathbf{x})$ implies that fluctuations of ξ on all scales, even of super–horizon size (where, due to the k^{-3} tail of the spectrum, ξ largely fluctuates) affect the present statistics of sub–horizon density perturbations which will then be non–Gaussian. The statistics of the peculiar gravitational potential so obtained can thus be used for building up the initial particle distribution in space and velocity through the Zel'dovich algorithm; this can then be used for evolving an N–body code [13]. Note that the Φ power–spectrum $P_\Phi(k)$ is simply related to that for the underlying Gaussian process ξ. In particular for the exponential potential model

$$P_\Phi(k) \propto (k/k_\star)^{-2\lambda^2/(2-\lambda^2)} k^{-3} \ln^{-1/3}(k/k_\star), \qquad (24)$$

with $k_\star = a_\star H_\star$. The logarithmic correction to the power–law spectral shape is the only effect of the t_1 dependence of $\xi_x(t_1)$ in the previous expressions; this is a general result which allows to treat ξ as being essentially time–independent in expressions such as (22) and (23). We can therefore conclude that inflationary models of the type considered here give curvature perturbations approximately described by the following form for the gravitational potential perturbation smoothed over the scale R

$$\Phi_R(\mathbf{x}) = \int d^3 y f_R(\mathbf{y} - \mathbf{x}) |\xi(\mathbf{y})|^\alpha + const, \qquad (25)$$

where $\xi(\mathbf{x})$ is now simplified to a Gaussian process with power–spectrum $P(k) = P_0 k^{-3}$, for $k_\star \leq k \leq k_{max}$, and $P(k) = 0$ elsewhere, k_{max} being a wave–number much larger than any mode of cosmological interest. The function f_R has Fourier transform

$$f_R(k) \propto W_R(k) T(k) k^{(n_p - 1)/2}, \qquad (26)$$

where $W_R(k)$ is a suitable low–pass filter which cuts off scales much smaller than R. Equations (25) and (26) allow to perform simulations of the non–Gaussian models described here, following standard methods (see [13]). It is important to note that the arbitrary additive constant in Eq.(25) has no observable effects.

Some authors have remarked (see, e.g., [7]) that the probability distribution for $\delta\varphi$ derived from the Fokker–Planck equation (8) cannot be used directly to yield density fluctuations for our observable patch of the inflated universe. As we have discussed here (see also [30],[9]) the obtaining of the actual density fluctuation field actually requires a more complete treatment taking into account the spatial variation of the quantum noise. This can be done by resorting to a functional Fokker–Planck equation or by using the Langevin equation (4) together with the η correlation function (7), as discussed so far. These authors have also

argued that one should somehow consider a different type of initial condition for Eq.(8), or a conditional probability that the coarse–grained field has a certain value φ at time t, i.e., on the scale R leaving the horizon at that time, given that it a had a particular value φ_0 at the time t_0, corresponding to the present horizon scale R_0. The time t_0 would typically correspond to about 60 e–foldings before the end of inflation; this is practically equivalent to starting to evolve the probability at t_0 from a delta function centered on a value $\varphi_0 \ll \varphi_\star$. The underlying idea is that one should somehow constrain the density field to be homogeneous on the horizon scale, since we cannot perceive inhomogeneities on scales much larger than R_0. It is clear that this condition prevents the system from entering the scaling regime, because one has too little time. Fluctuations in the gravitational potential, however, still happen to be small, because the dispersion had little time to grow. In this case one can easily show that for most values of φ_0 the fluctuations are essentially Gaussian, with the exception of very high fluctuations, rare events which do not perceive the constraint. Also, for a few values of the constraint φ_0, the conditional probability is highly non–Gaussian and similar to the unconstrained one: these rare values of φ_0 give in fact the dominant weight in the unconstrained probability, as a consequence of the intermittency property, referred above. We believe, however, that the constraint imposed by the value of the coarse–grained field φ_0 is unphysical, because it ignores the arbitrary additive constant which enters the definition of the gravitational potential. Lucchin, Matarrese and Ortolan [9] therefore consider a different quantity, which is not affected by this indeterminacy: the gradient of the gravitational potential $\mathbf{\Psi}$, on a given scale R, this being proportional to the peculiar velocity field. For simplicity we consider the quantity $\Delta\Psi \equiv [\mathbf{\Psi}(\mathbf{x}) - \mathbf{\Psi}_0(\mathbf{x})] \cdot \mathbf{n}$, with $\mathbf{\Psi}_0$ the same quantity evaluated on the scale R_0. This quantity is proportional to the peculiar velocity, projected along the direction n, measured by an observer placed in x in the local rest frame set by the cosmic background radiation. Both the probability for $\Delta\Psi$ and the one for the same quantity conditioned by the value of $\Psi_0 \equiv \mathbf{\Psi}_0 \cdot \mathbf{n}$ are highly non–Gaussian, for essentially all values of the constraint. These non–Gaussian distributions are characterized by power–law tails (instead of exponential ones, like in the Gaussian case), namely

$$\mathcal{P}(\Delta\Psi) \sim |\Delta\Psi|^{-\gamma}, \qquad \mathcal{P}(\Delta\Psi|\Psi_0) \sim |\Delta\Psi|^{-(1+\gamma)}, \tag{27}$$

with $\gamma = 1 + 1/(1 - \alpha)$, which typically imply diverging moments. [More details will be given in the paper by Lucchin, Matarrese and Ortolan [9].] The physical consequence of this fact is that high peculiar velocities on large scales are much more likely than for a Gaussian field with the same power–spectrum. This qualitative conclusion is confirmed by numerical simulations of initially non–Gaussian distributions for the gravitational potential [13].

CONCLUSIONS

The stochastic approach allows to study the dynamics of inflation on scales larger than the Hubble radius. It accounts for the generation of large–scale classical fluctuations from quantum oscillations inside the horizon, the effect of non–linearities on the evolution of inflaton perturbations and the back–reaction of matter fluctuations on the background geometry.

If the spatial dependence of the fine–grained correlation function is kept, the whole spatial pattern of the scalar field fluctuations is known: from this the statistical distribution of the gravitational potential fluctuation field can be completely reconstructed. A generic feature of models leading to chaotic or to power–law inflation is the occurrence of a scaling regime where the coarse–grained distribution sharply peaks around the classical solution. In this regime the distribution is non–Gaussian and scale–invariant, due to the dominance of the inflaton non–linearities over the diffusion caused by quantum fluctuations. The deviation from the Gaussian behaviour does not depend on the strength λ of the non–linearity, on time (scale), and on the initial condition, provided the latter permits the system to enter the scaling regime. Because the probability is peaked, its bulk properties may be described by a Gaussian centered on $\varphi_{cl}(t)$ with suitable dispersion, but this approximation would fail in estimating the likelihood of rare events for which the actual distribution is required.

The scale–invariance of the inflaton is reproduced by the peculiar gravitational potential field, during its linear evolution. As a consequence, the N–point mass correlation functions will obey Eq.(2), with a spectral index n_p determined by the inflationary parameters; in the exponential potential model, for instance, $n_p = 1 - 2\lambda^2/(2 - \lambda^2) \leq 1$ (up to negligible logarithmic corrections). The non–Gaussian scaling invariance and the multiplicative character of the primordial perturbation field will affect the properties of the universe on large scales, in particular the probability for the occurrence of rare events, such as high peculiar velocities, large empty regions, long filaments and *great attractors*. Moreover, one should expect interesting consequences for the statistics of the cosmic microwave background anisotropies.

ACKNOWLEDGEMENTS

REFERENCES

1. Kofman, L.K. and Linde, A., 1987, Nucl. Phys. **B282**, 555.
2. Allen, T.J., Grinstein, B. and Wise, M.B., 1987, Phys. Lett. **197B**, 66.
3. Kofman, L.K. and Pogosyan, D.Yu., 1989, Phys. Lett. **214B**, 508.
4. Salopek, D.S., Bond, J.R. and Bardeen, J.M., 1989, Phys. Rev. **D40**, 1753.
5. Ortolan, A., Lucchin, F. and Matarrese, S., 1988, Phys. Rev. **D38**, 465.
6. Matarrese, S., Ortolan, A. and Lucchin, F., 1989, Phys. Rev. **D40**, 290.
7. Kofman, L.K., Blumenthal, G.R., Hodges, H. and Primack, J.R., 1990, in *Proc. Workshop on Large Scale Structures and Peculiar Motions in the Universe*, Rio de Janeiro, May 1989, Latham, D.W., and da Costa, L.N. eds.; *in press*.

8. Barrow, J.D. and Coles, P., 1990, Mon. Not. R. astr. Soc. *in press.*

9. Lucchin, F., Matarrese, S. and Ortolan, A., 1990, *in preparation.*

10. Linde, A.D., 1983, Phys. Lett. **129B**, 177.

11. Abbott, L.F. and Wise, M.B., 1984, Nucl. Phys. **B244**, 541.

12. Lucchin, F. and Matarrese, S., 1985, Phys. Rev. **D32**, 1316.

13. Lucchin, F., Matarrese, S., Messina A. and Moscardini L., 1990, *in preparation.*

14. Otto, S., Politzer, H.D., Preskill, J. and Wise, M.B., 1986, Ap. J. **304**, 62.

15. Lucchin, F., Matarrese, S. and Vittorio, N., 1988, Ap. J. (Letters) **330**, L21.

16. Zel'dovich, Ya.B., 1970, Astr. Astrophys. **5**, 84.

17. Peebles, P.J.E., 1983, Ap. J. **274**, 1.

18. Zel'dovich, Ya.B., Molchanov, S.A., Ruzmaikin, A.A. and Sokolov, D.D., 1987, Sov. Phys. Usp. **30**, 5.

19. Hodges, H., 1990, Phys. Rev. Lett. **64**, 1080.

20. Mollerach, S., Lucchin, F., Matarrese, S., and Ortolan, A., 1990, *in preparation.*

21. Starobinskii, A.A., 1986, in *Field Theory, Quantum Gravity and Strings*, ed. by H.J. de Vega and N. Sanchez (Lecture Notes in Physics, **246**) (Springer–Verlag, Berlin).

22. Bardeen, J.M. and Bublik, G.J., 1987, Class. Quantum Grav. **4**, 573.

23. Goncharov, A.S., Linde A.D. and Mukhanov, V.F., 1987, Int. J. Mod. Phys. **A2**, 561.

24. Rey, S.-Y., 1987, Nucl. Phys. **B284**, 706.

25. Mijic, M., 1990, *Sissa preprint, 18 A.*

26. Suzuki, M., 1981, Adv. Chem. Phys. **46**, 195.

27. Sasaki, M., Nambu, Y. and Nakao, K., 1988, Nucl. Phys. **308**, 868.

28. Graziani, F.R., 1989, Phys. Rev. **D39**, 3630.

29. Hodges, H., 1989, Phys. Rev. **D39**, 3568.

30. Matarrese, S., Lucchin, F. and Ortolan, A., 1990, in *Proc. Workshop on Large Scale Structures and Peculiar Motions in the Universe*, Rio de Janeiro, May 1989, Latham, D.W., and da Costa, L.N. eds.; *in press.*

INFLATION AND THE BARYON ISOCURVATURE MODEL

Silvia Mollerach
International School for Advanced Studies
Strada Costiera 11, 34014 Trieste, Italy

ABSTRACT

It has been proposed that a scalar field present during the inflationary era in addition to the inflaton, which decays into thermal radiation after the baryogenesis can produce spatial fluctuations in the initially smooth entropy per baryon ratio. These perturbations were hoped to be of the isocurvature type and it was expected that they may be useful to explain some observed features of the large scale structure. However, a detailed study of the generation of perturbations in this two-field inflationary model shows that the resulting fluctuations are not of the isocurvature type, but that the entropy perturbation induces a curvature fluctuation during the evolution of the perturbations outside the Hubble radius, which is larger than the entropic one. Thus, this model is not a good candidate to provide the initial conditions needed in the baryon isocurvature model.

INTRODUCTION

In the simplest model, density fluctuations originated during inflation are of the adiabatic type [1]. The reason is that when the inflaton decays, reheating the universe, the fluctuations in all the decay components follow the original inflaton fluctuations. Baryogenesis occurs after this, thus the resulting entropy per baryon is constant in space. However, it has been argued that this is not the only possibility. Isocurvature perturbations can also be produced provided that there is another scalar field present during inflation besides the inflaton. This idea has first been proposed in relation to the axion field [2], but has then been generalized to other weakly interacting scalar fields [3]. Further, it has recently been pointed out by Peebles [4] that if the second scalar field decays into radiation after the baryogenesis, the density fluctuations associated to it give rise to fluctuations in the previously smooth entropy per baryon ratio, which are absent in other models that do not modify the baryogenesis mechanism. Another model for the origin of baryon isocurvature perturbations has been proposed recently [5], based in a new model for baryogenesis [6], in which the baryon number per entropy originated is a function of the space point.

On the other hand, isocurvature perturbations in phenomenological models have attracted much attention recently, as some controversial points associated to the standard cold dark matter adiabatic perturbation model [7] have arisen. The problems are essentially that in this model the fluctuations in the mass distribution are anticorrelated on scales larger than $\sim 50 - 100h^{-1}Mpc$, which seems to be inconsistent with the observation of large-scale velocity fields and structures in the galaxy distribution [8] and that the epoch of galaxy formation seems to occur too late. This calls attention to alternative models for galaxy formation. It has been argued that in models with baryon isocurvature initial perturbations [4] galaxies can form early ($z \sim 30$), and mass fluctuations on the scale λ_J can drive large scale velocity fields [9]. However, detailed studies of the cosmic background radiation anisotropies impose stringent bounds on isocurvature models and may even rule them out unless sufficient reionization occurs after recombination [10]. In the phenomenological models it is taken as initial conditions that the total energy density is homogeneous in space but entropy is inhomogeneous (which means that the ratio of the densities corresponding to the different components is perturbed). These initial conditions are imposed during the radiation dominated era and the following evolution is then computed in the multicomponent universe [11].

It has recently been shown [12] that the model with an extra scalar field present during

inflation which decays into radiation after baryogenesis [4], does not actually provide the initial conditions needed in the baryon isocurvature model as was expected. The reason is that, even if the relative fluctuations in the energy density can be much smaller for the inflaton than for the other scalar field, which means that the fluctuations are initially of the isocurvature type, the entropy perturbations act as source for the total density perturbations. Curvature fluctuations induced by this source become the dominant ones at the radiation dominated era. Other discussions related to this problem can be found in ref.[13,14]. A similar analysis of other models proposed in the literature to originate isocurvature perturbations shows that the case in which the second scalar field remains as a dark matter component up to the present, the axion model and the spontaneous baryogenesis one provide instead good initial conditions for phenomenological isocurvature perturbations [15].

TWO NON-INTERACTING SCALAR FIELDS

It has been pointed out in refs.[3] that isocurvature or isothermal fluctuations are quite generic within the inflationary scenario. In fact, many types of scalar fields appear in elementary particle theories and some of them interact very weakly with the rest of the particles. If some of these fields were present during inflation (contributing much less than the inflaton to the total energy density), since after the reheating their energy density decreases more slowly than that of the radiation produced by the inflaton decay, they can give an important contribution to the total density at a later period. We will analyse here the model proposed by Peebles, in which the additional scalar field χ decays into thermal radiation after becoming the dominant component, giving rise to fluctuations in the number of photons per baryons. The evolution of the background energy density of both components can be seen in the figure.

In order to study the resulting fluctuations in this model there are two main steps. The first is to compute the perturbations in each component at the Hubble radius crossing during the inflationary era originated by the quantum fluctuations of both fields, and the second is to follow their classical evolution afterwards. This requires to study the evolution of the perturbations in the multicomponent system composed by the inflaton, its decay products, the other scalar field, and eventually the products of its decay (for example in the Peebles model analysed here, the scalar field decays into radiation). This study is simplified if we consider one component (denoted by α) formed by the inflaton ϕ, and the radiation and baryons in which it decays ($\phi+\mathrm{rad}_\phi+\mathrm{bar}_\phi$) and another component (denoted by β) by the other scalar field and its decay products ($\chi+\mathrm{rad}_\chi$). The evolution of the perturbations in the two-component system is studied using a formalism developed by Kodama and Sasaki [16].
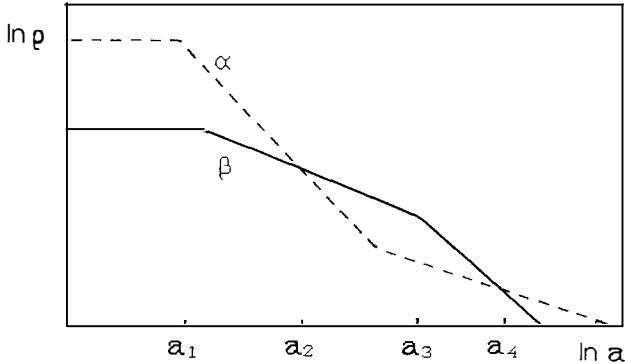
Evolution of the background model. The dashed line corresponds to ρ_α and the continuous one to ρ_β. a_1 corresponds to the scale factor at the end of the inflation, a_2 to that when the χ field becomes dominant, a_3 to that when χ decays into radiation and a_4 to that when the universe becomes matter dominated.

The most convenient variables to study this system are the comoving perturbation of the total energy density and velocity, Δ and V, the entropy perturbation $S_{\alpha\beta}$, and the perturbation in the relative velocity between the components $V_{\alpha\beta}$ defined by

$$S_{\alpha\beta} \equiv \frac{\delta_\alpha}{1 + w_\alpha} - \frac{\delta_\beta}{1 + w_\beta}, \qquad V_{\alpha\beta} \equiv v_\alpha - v_\beta, \tag{1}$$

where $\delta_\alpha \equiv \delta\rho_\alpha/\rho_\alpha$ and $w_\alpha \equiv p_\alpha/\rho_\alpha$, in terms of the energy density ρ and the pressure p. $S_{\alpha\beta}$ and $V_{\alpha\beta}$ are gauge invariant variables. Δ is related to the Bardeen potential Φ_H [14] by $\Phi_H \equiv (3/2)(k/aH)^2 \Delta$.

Adiabatic fluctuations are characterized by having $S_{\alpha\beta} = 0$, which means that all the components fluctuate in the same way. Isocurvature fluctuations instead correspond to relative fluctuations between the different components such that the total curvature is left invariant ($\Delta \ll S_{\alpha\beta}$).

Quantum fluctuations of the scalar fields lead to fluctuations in the energy density associated to each field. However, the fact that the contribution of χ to the total energy density is small during inflation ensures that their fluctuations do not perturb the total energy density too much (the major contribution to the total ρ perturbations is given by the inflaton ϕ fluctuations). But it can be seen [12] that the fluctuations in the entropy density are dominated by the second scalar field χ and are typically larger than the fluctuations in the total energy

density. The result is that at Hubble radius crossing

$$\Delta\Big|_{H} \simeq -\frac{\dot{\phi}\, D\phi}{3H}\Big|_{H}, \qquad V_{\alpha\beta}\Big|_{H} \simeq -\frac{H\, D\chi}{\dot{\chi}}\Big|_{H},$$

$$V\Big|_{H} \simeq \frac{H\, D\phi}{\dot{\phi}}\Big|_{H}, \qquad S_{\alpha\beta}\Big|_{H} \simeq 4\frac{H\, D\chi}{\dot{\chi}}\Big|_{H}. \tag{2}$$

(Units are chosen so that $c = 8\pi G = \hbar = 1$.) Fluctuations in both fields can be estimated for the purpose of our study by the amplitude of quantum fluctuations of a massless scalar field in the de Sitter spacetime

$$D\phi \sim D\chi \sim \frac{H}{\sqrt{2\pi}k^{3/2}}. \tag{3}$$

Note that expressions (2) imply that the fluctuations are of the isocurvature type when they leave the Hubble radius during the inflationary era, as $S_{\alpha\beta}\big|_{H} \gg \Delta\big|_{H}$ (during inflation $\dot{\phi}, \dot{\chi} \ll H$).

The following evolution of the perturbations is determined by the set of equations

$$\frac{d^2\Delta}{da^2} + \left(\frac{3}{2} - \frac{15}{2}w + 3c_s^2\right)\frac{1}{a}\frac{d\Delta}{da} + \left(-\frac{3}{2} + 3c_s^2 - 12w + \frac{9}{2}w^2 + c_s^2\left(\frac{k}{aH}\right)^2\right)\frac{\Delta}{a^2} =$$

$$\left(\frac{k}{aH}\right)^2 \frac{1}{a^2}\left(\frac{h_\alpha h_\beta}{h\rho}(c_\alpha^2 - c_\beta^2)S_{\alpha\beta} + \frac{p_\alpha\eta_\alpha + p_\beta\eta_\beta}{\rho}\right), \tag{4.a}$$

$$\frac{dS_{\alpha\beta}}{da} = -\frac{k}{aH}\frac{V_{\alpha\beta}}{a} - 3\left(\frac{w_\alpha}{1+w_\alpha}\frac{\eta_\alpha}{a} - \frac{w_\beta}{1+w_\beta}\frac{\eta_\beta}{a}\right), \tag{4.b}$$

$$\frac{dV_{\alpha\beta}}{da} + \left(1 - 3(c_{s\alpha}^2\frac{h_\beta}{h} + c_{s\beta}^2\frac{h_\alpha}{h})\right)\frac{V_{\alpha\beta}}{a} = \frac{k}{aH}\left(c_{s\alpha}^2\frac{h_\beta}{h} + c_{s\beta}^2\frac{h_\alpha}{h}\right)\frac{S_{\alpha\beta}}{a} +$$

$$\frac{k}{aH}\frac{(c_{s\alpha}^2 - c_{s\beta}^2)}{1+w}\frac{\Delta}{a} + \frac{k}{aH}\frac{1}{a}\left(\frac{w_\alpha\eta_\alpha}{1+w_\alpha} - \frac{w_\beta\eta_\beta}{1+w_\beta}\right). \tag{4.c}$$

where $p_\alpha\eta_\alpha = \delta p_\alpha - c_{s\alpha}^2\delta\rho_\alpha$, $h_\alpha = \rho_\alpha + p_\alpha$ and $c_{s\alpha}^2 = \dot{p}_\alpha/\dot{\rho}_\alpha$. These equations allow us to follow the evolution of the perturbed variables during the different periods determined by the equations of state of the two components. They can be solved for each period with w_α and w_β approximately constant. The initial conditions are given by (2) and the matching between periods is made by imposing continuity of all the fluctuation variables. Note that the right hand side of eq. (4.a) vanishes when we are dealing with adiabatic perturbations. This source term, present in the non-adiabatic multicomponent case, is due to the extra contribution to the total pressure perturbations given by the entropic perturbation $S_{\alpha\beta}$ and the non-adiabatic pressure perturbation of the individual components η_α (this term can be present, for example, when one of the components corresponds to a scalar field, because both

the energy and the pressure perturbations are determined by the scalar field fluctuations and there is no extra freedom to fix some relation between them). In the case of isocurvature perturbations, we have that initially Δ is very small, but $S_{\alpha\beta}$ is large, so its effects cannot be neglected, even outside the Hubble radius. Their quantitative importance can be computed solving the evolution equation for $S_{\alpha\beta}$ and replacing this solution in the source term of the evolution equation for Δ. The presence of $S_{\alpha\beta}$ can make grow a perturbation Δ even if the adiabatic modes are initially set to zero. When a change in the equation of state occurs, the modes get mixed and in a following period the adiabatic modes can become excited. It has been shown in ref. [12] that in the model in which there is an additional scalar field χ present besides the inflaton, which decays into radiation after the baryogenesis, an initially pure isocurvature fluctuation at Hubble radius crossing ($H1$) during inflation,

$$\Delta\Big|_{H1} \sim 0 \qquad S_{\alpha\beta}\Big|_{H1} \neq 0, \tag{5}$$

produces, by the time the perturbations reenter the Hubble radius ($H2$) an adiabatic mode with amplitude

$$\Delta\Big|_{H2} \sim \mathcal{O}(10^2)\, S_{\alpha\beta}\Big|_{H2}. \tag{6}$$

Thus, the perturbations are no longer of the isocurvature type by this time and they do not provide the initial conditions required by the phenomenological baryon isocurvature model. The reason is that in the phenomenological model the isocurvature initial condition ($\Delta = 0, S_{\alpha\beta} \neq 0$) is imposed at the radiation dominated epoch before baryons become the dominant component, instead in the model proposed in ref. [4] the adiabatic mode responsible for the large density fluctuation at $H2$ (eq. (6)) has become significant much earlier, at the epoch when the universe was dominated by the oscillating χ field.

It can be seen, through computations similar to those in [12], that the growth of total density fluctuations outside the Hubble radius is quantitatively important also in the models in which the second scalar field does not decay into radiation, contributing now to the dark matter. However, this is a good model for phenomenological isocurvature perturbations. The point is that the isocurvature initial conditions are imposed in the radiation dominated era, just after the decay of the inflaton, and this holds as a very good approximation in this model. The growth of the total density perturbation becomes significant only when the χ contribution to the total density begins to be important, which happens later. Thus the resulting total density perturbations are correctly described within the phenomenological models. Furthermore, much interest have been concentrated in these models because of the possibility of getting a spectrum of fluctuations different from the scale invariant one [3,17].

PSEUDO-GOLDSTONE BOSON

The other kind of model proposed for the origin of isocurvature perturbations consists of a pseudo-goldstone boson which is non-massive during inflation and acquires a mass in a later period. In these models the entropy perturbations are generated after the end of inflation and thus the isocurvature condition is valid during the radiation dominated epoch. The best known example is the axion. At very large temperatures, the axion potential is essentially flat and it acquires a small mass through QCD non-perturbative effects at approximately $T \sim 1 GeV$. As the axion interacts very weakly with the rest of the matter, it oscillates around the minimum of the potential during the following evolution of the universe, behaving like non-relativistic particles (cold dark matter). The origin of the density fluctuations in this model has been widely studied [2]. The idea is that quantum fluctuations of the massless axion during inflation give rise to spatial inhomogeneities of the axion distribution and when the potential of the axion becomes non-trivial due to QCD instanton effects, these inhomogeneities are translated into fluctuations of the axion energy density. If this process does not alter the total energy density too much, the resulting fluctuations are of the isocurvature type. As this condition holds during the radiation dominated era, the adiabatic modes are not excited in this period, they only appear later, when the universe becomes axion dominated.

Finally, let us consider the spontaneous baryogenesis model. The mechanism for the generation of energy density fluctuations is quite similar to that in the axion model. In this case, the fluctuations produced are in the number of baryons per entropy. They arise from B violating interactions of a pseudo-goldstone boson (the ilion) as it relaxes to the true minimum of its potential. This occurs after the end of inflation. Thus, the isocurvature condition ($\Delta = 0$ and $S_{\alpha\beta} \neq 0$) is imposed during the radiation dominated era and, as in the axion case, the adiabatic modes are not excited in this period. This fact makes the resulting fluctuations a possible model for the phenomenological baryon isocurvature fluctuations.

SUMMARY

In summary, we have analysed the different models proposed for the origin of isocurvature fluctuations. In particular, we have studied if they can provide the initial conditions needed for the phenomenological isocurvature model, which means that the growing adiabatic mode is not excited in the radiation dominated era. The result is that in the models with an extra weakly interacting field present during inflation, in the case in which it decays into radiation after inflation, the original entropy perturbations induce a large adiabatic mode by

40

the radiation dominated period (when initial conditions are set in phenomenological models).
Instead, the case in which the extra scalar field remains as a dark matter component up to the
present epoch, the axions and the model of spontaneous baryogenesis are possible candidates
to originate this kind of fluctuations.

References

[1] See for e.g.: S. Hawking, Phys. Lett. B115 (1982) 295, A. H. Guth and S.-Y. Pi,
     Phys. Rev. Lett. 49 (1982) 1110, A. D. Linde, Phys. Lett. B116 (1982) 335, A. A.
     Starobinski, Phys. Lett. B117 (1982) 175, A. H. Guth and S.-Y. Pi, Phys. Rev. D32
     (1985) 1899, J. M. Bardeen , P. J. Steinhardt and M. S. Turner, Phys. Rev. D28 (1983)
     697, V. F. Mukhanov, JETP 41 (1985) 493.

[2] M. Axenides, R. H. Brandenberger and M. S. Turner , Phys. Lett. B126 (1983) 178, A.
     D. Linde, Phys. Lett. B158 (1985) 375, D. Seckel and M. S. Turner, Phys. Rev. D32
     (1985) 3178.

[3] A. D. Linde, JETP Lett. 40 (1984) 1333, L. A. Kofman, Phys. Lett. B173 (1986) 400,
     L. A. Kofman and A. D. Linde, Nucl. Phys. B282 (1987) 555.

[4] Peebles P.J.E., in: Large scales structures and motions in the universe (Trieste, April
     1988), ed. Mezzetti, Giuricin, Mardirossian and Ramella (Kluwer Academic Plublishers,
     Dordrecht, 1989) pag. 119.

[5] M. S. Turner, A. G. Cohen and D. B. Kaplan, Phys. Lett. B216 (1989) 20.

[6] A. G. Cohen and D. B. Kaplan, Nucl. Phys. B308 (1988) 913.

[7] see e.g. C. S. Frenk, S. D. M. White, M. Davis and G. Efstathiou, Astrophys. J. 327
     (1988) 507 and references therein.

[8] A. Dressler et al., Astrophys. J. Lett. 313 (1987) L37, M. P. Haynes and R. Giovanelli,
     Astrophys. J. Lett. 306 (1986) L55.

[9] C. J. Hogan and N. Kaiser, Astrophys. J. 274 (1983) 7, P. J. E. Peebles, Nature 327
     (1987) 210.

[10] P. J. E. Peebles, Astrophys. J. 315 (1987) L73, G. Efstathiou and J. R. Bond, Month.
     Not. of R. Ast. Soc. L27 (1987) 33p, K. M. Gorski and J. Silk, preprint (1989), N.
     Gouda, M. Sasaki and Y. Suto, Astrophys. J. 341 (1989) 557.

[11] H. Kodama and M. Sasaki, Int. J. Mod. Phys. A1 (1986) 265, H. Kodama and M.
     Sasaki, Int. J. Mod. Phys. A2 (1987) 491, G. Esfstathiou and J. R. Bond, Mon. Not.
     R. ast. Soc. 218 (1986) 103.

[12] S. Mollerach, (to be published in Phys. Rev. D)

[13] W. H. Press and E. T. Vishniac, Astrophys. J. 239 (1980) 1, V. N. Lukash, in: Large
     scales structures and motions in the universe (Trieste, 1988), ed. Mezzetti, Giuricin,
     Mardirossian and Ramella (Kluwer Academic Plublishers, Dordrecht, 1989) p. 139.

[14] J. M. Bardeen, Phys. Rev. D22 (1980) 1882.

[15] S. Mollerach, (to be published in Phys. Lett. B)

[16] H. Kodama and M. Sasaki, Prog. of Theor. Phys. Supp. 78 (1984).

[17] D. S. Salopek, J. R. Bond and J. M. Bardeen, Phys. Rev. D40 (1989) 1753, L. A.
     Kofman and D. Yu. Pogosyan Phys. Lett. B214 (1988) 508.

PHASE TRANSITION
AND NUCLEOSYNTHESIS

BARYOGENESIS IN THE UNIVERSE AND BARYONIC CHARGE CONDENSATE

A.D.DOLGOV
Institute of Theoretical and Experimental Physics
117259 Moscow, USSR

ABSTRACT
    After a brief review of the basic ideas of the baryogenesis a model of low temperature generation of the baryon asymmetry of the Universe is discussed. The model is based on the assumption of baryonic charge condensate formation during inflationary stage. It is shown that the results crucially depend upon the rate of the particle production by the condensate. The model possesses two unusual features: leptonic charge asymmetry may be large as compared to the baryonic one and the characteristic scale of baryonic or leptonic charge density variation may be relatively small. The latter can be traced by the large scale structure of the Universe or by the helium-4 distribution respectively.

Baryon or charge asymmetry of the Universe is now well explained by the laws of fundamental physics. In principle one can express the observed ratio of number densities

$$\beta = (N_{\delta} - \overline{N}_{D})/N_{\tau} = 3 \cdot 10^{-10} \tag{1}$$

through the particle masses and the coupling constants of their interactions. Unfortunately there are too many different theoretical models which give the results in reasonable agreement with (1) and at the moment we don't know what possibility is realized in Nature.

Three sacred principles of baryogenesis formulated by Sakharov [1], i.e.

1. Baryonic charge nonconservation;

2. Breaking of charge (C and CP) symmetry, and

3. Deviation from thermal equlibrium

present sufficient conditions for generation of charge asymmetry asymmetry in initially charge symmetric state. All these postulates seem to be well established now either theoretically or experimentally but, what became known relatively recently, none of them is necessary.

Baryonic charge nonconservation which was the weakest point originally now is quite respectable , the minor shortcoming being that proton decay or any other manifestations of nonconservation of baryons in direct experiments are not yet observed. Almost all unification models predict a strong B-nonconservation at the unification scale m_{GUT}. More interesting is that the electroweak theory in which B is conserved at the classical level predicts B-nonconservation due to quantum anomaly [2]. The effect is tiny ($\sim exp(-2\pi/\alpha)$) at low energy but it is possible that at high temperatures or energies it may be

not suppressed. In that case electroweak reactions can drastically change baryonic charge density in the Universe at the temperatures about 1 TeV [3].

A weighty argument in favour of B-nonconservation presents inflation. One knows that in order to solve the problems of flatness, horizon, homogeneity and isotropy in the Friedman cosmology there must exist a stage of the exponentially fast expansion, $a \sim exp(Ht)$ with duration not smaller than $t \cdot H = 70$. This expansion regime is realized if the energy density in the Universe is constant. On the other hand if B is conserved the energy density of baryons changes as a^{-4}. It follows from eq.(1) that relative baryonic energy density at $T > 1$ GeV is $\rho_B/\rho_{tot} > 10^{-10}$. Thus for conserved B inflation could not last longer than six Hubble times, $t \cdot H < 6$, otherwise the condition $\rho_{tot} = const$ can not be fulfilled.

At a first glance baryonic charge nonconservation is obligatory for a dynamical generation of baryon asymmetry. Black hole evaporation however presents a counterexample to that. It could give rise to an exsess of baryons over antibaryons even if B is strictly conserved in particle interactions[4]. A simple toy model of this phenomenon is the following. Let A be a meson emitted by the black hole and let A decays into channels A --> $L\bar{H}$, $H\bar{L}$ where H and L are heavy and light baryons respectively, and $\bar{H}$ and $\bar{L}$ are their antiparticles. If the branching ratio of A --> $L\bar{H}$ is larger than that of A --> $\bar{L}H$ then the black hole should predominantly emit baryons because the back capture of a heavy particle by a black hole is larger than that of a light one.

The necessity of nonequilibrium for baryosynthesis can be

easily seen because the equilibrium number densities are de-
termined by mass and chemical potential of the particle,

$$N = \int \frac{d^3p}{(2\pi)^3} \left[exp \left(\frac{\sqrt{p^2+m^2}}{T} - \frac{\mu}{T} \right) \pm 1 \right]^{-1}$$

Because of CPT-theorem masses of particle and antiparticle
must be equal, $m=\bar{m}$, and chemical potentials μ and $\bar{\mu}$ must va-
nish if the corresponding charge is not conserved. So in
equilibrium number densities of particles and antiparticles
with nonconserved charge must be the same. This is not true
however in the nonstationary case if there exists the interac-
tion of the type[5]:

$$L_I = \frac{1}{f} \partial_\mu \varphi \cdot J_\mu^B \tag{2}$$

where f is a parameter with dimension of mass, φ is a scalar
field, and J_μ^B is nonconserved baryonic current. If φ is a
classical homogeneous field, $\varphi=\varphi(t)$, with time dependence de-
termined by the Universe expansion, interaction (2) gives rise
to the baryonic chemical potential

$$\mu_B = \frac{\dot{\varphi}}{f} \tag{3}$$

which is nonzero even in thermal equilibrium.

The same model permits baryosynthesis without explicite
breaking of charge symmetry but I don't dwell on it here and
return to it in what follows because the model I am going to
discuss in this talk possesses this property.

In hystorically first models of baryogenesis[1,6] it was
assumed that it proceeds at very high temperatures $T=m_{GUT}=10^{15}$
GeV. In this way it proves possible ensure the observed proton
stability (or, better to say, a very long life time) with suf-
ficiently fast B-nonconserving processes. These models seem to

be out of question, or maybe out of fashion, now. First the reheating temperature after inflation T_R is most probably smaller than T_{out} and second, as is well known, if $T_R > 10^{12}$ GeV one encounters the problem of overabundant relic graviti- nos[7]. Fortunately modern particle theory permits low tempera- ture baryoproduction either at electroweak-scale $T_{ew} = O(1TeV)$ or even at smaller temperatures $T \approx 100$ GeV. The last possibili- ty is the topic of this talk.

A low temperature scenario of baryogenesis was proposed in ref.[8]. If was assumed that there exists a scalar field χ with nonzero and nonconserved baryonic charge and vanishing all conserved charges. Such fields exist in a class of super- symmetric models. During inflationary stage such a field gene- rically develops a vacuum condensate if the field mass m_χ is small in comparison with the Hubble parameter H_I and the sel- finteraction is weak. Indeed quantum fluctuations of a free massless scalar field φ in the de Sitter background is known[7] to rise as

$$\langle \varphi_0^2 \rangle = \frac{H^3 t}{4\pi^2} \qquad (4)$$

The increase terminates when the field potential energy $U(\phi)$ becomes of the order of kinetic energy of quantum fluctu- ations $E_K(\phi)=1/2(H/2\pi)^4(\partial\phi)^2$. If ϕ is a massive free field, i.e. $U(\phi)=m^2\phi^2/2$, then [10]

$$\langle \varphi_m^2 \rangle \rightarrow \frac{3H^4}{8\pi^2 m^2} \qquad (5)$$

Using the virial considerations one can evaluate the limiting value of $\langle \phi^2 \rangle$ for ϕ-selfintraction of the form $U(\phi)= \lambda\phi^4$. In this case

$$\langle \varphi_\lambda^2 \rangle = \frac{H^2}{\sqrt{\lambda}} \cdot const \tag{6}$$

where const=O(1).

Since any wavelength exponentially increases during inflation the small scale quantum flucluations of φ stretch up and quantum field turns into a classic one. In that sence one could speak about the field condensate on macroscopically large scale.

If a scalar field has a conserved charge the condensate is not developed because the density of any conserved charge must go down as $a^{-3}(t)$, where $a(t)$ is the scale factor (during inflation $a \sim exp(H_I t)$).

Since χ has a nonconserved baryonic charge its density during inflation should be of the order of H_I^3 (by dimensional grounds). When inflation stops and the Hubble parameter becomes smaller than m_χ the χ-condensate decays into massless quarks and leptons. Baryonic charge in the model considered is conserved in these decays and thus the vacuum condensate of baryonic charge is transformed into baryonic charge of quarks giving rise to the baryon asymmetry of the Universe.

In the original version of the model [9] the value of β was of the order of unity, i.e. ten orders of magnitude larger than that given by observations (1). There were a few attempts to make the model consistent with observations [11]. However as we see in what follows, the situation is quite opposite: the model naturally gives very small values of β and only for a rather special values of the coupling constants one may expect $\beta=O(1)$ [12].

Let consider the following toy model

$$V(\chi) = m^2 |\chi|^2 + \frac{1}{2} \lambda_1 |\chi|^4 + \frac{1}{4} \lambda_2 (\chi^4 + \chi^{*4}) \tag{7}$$

Due to the last term the potential is not invariant with respect to the phase rotation $\chi \longrightarrow \chi \, exp(id)$ and the baryonic current of χ is not conserved

$$\partial_\mu j^{B}_\mu \equiv \partial_\mu (i \chi^* \overset{\leftrightarrow}{\partial_\mu} \chi) = i \lambda_2 (\chi^{*4} - \chi^4) \tag{8}$$

It is very convenient to have in mind the following mechanical analogy. The equations satisfied by homogeneous field $\chi(t) = \chi_1(t) + i \chi_2(t)$ in the Robertson–Walker metric:

$$\ddot{\chi_1} + 3H\dot{\chi_1} + [m^2 + \chi_1^2(\lambda_1 + \lambda_2) + \chi_2^2(\lambda_1 - 3\lambda_2)] \chi_1 = 0,$$

$$\ddot{\chi_2} + 3H\dot{\chi_2} + [m^2 + \chi_2^2(\lambda_1 + \lambda_2) + \chi_1^2(\lambda_1 - 3\lambda_2)] \chi_2 = 0 \tag{9}$$

are equivalent to the mechanical equation of motion of a point particle in the two-dimensional plane (χ_1, χ_2) with potential $V(\chi)$ and the friction coefficient 3H. The baryonic charge in this language is the angular momentum of the particle and its nonconservation is enforced by the nonsphericity of the potential.

It was assumed in refs.[6,7] that potential $V(\chi)$ has flat directions in which V does not rise with increasing χ . In our model it is realized if e.g. $\lambda_1 = -\lambda_2 = \lambda > 0$. If this is the case the directions $\chi_2 = 0$ and $\chi_1 = 0$ are flat and χ_1 or χ_2 respectively can rise along one or other flat direction in accordance with eqs.(4) and (5). The variation of χ in the orthogonal direction in governed by the potential $4\lambda \chi_1^2 \chi_2^2$. This means that the effective mass of, say, χ_2 in the state, where classical field χ_1 is developed, goes up with increasing

50

$$\langle\chi_1\rangle \text{ as}$$

$$m(\chi_2) = 2\sqrt{\lambda}\langle\chi_1\rangle = \min\left(\frac{H}{2\pi}\sqrt{Ht}\; ;\; \frac{1}{2\pi}\sqrt{\frac{3}{2}}\,\frac{H}{m_\chi}\right) \tag{10}$$

The walls in the valley become steeper the further away is the point from the origin.

The baryonic charge density in this state is

$$\sqrt{\langle N_B^2\rangle} = \sqrt{\langle j_0^{B\,2}\rangle} = -2\dot{\chi}_2\chi_1 = O\left(H_I^2\chi_{10}\right) \tag{11}$$

One can easily see from the mechanical picture that the angular momentum periodically changes sign, so that $\langle N_B\rangle = 0$. These oscillations are damped by the Universe expansion (Hubble friction) and by the particle production by the time dependent field χ_2. The first mechanism is relatively unessential. If the particle production is not taken into account the field χ goes down to the origin with large baryonic charge which ultimately transforms into baryonic charge of quarks. This takes place when χ becomes small enough so the term $m^2|\chi|^2$ in the potential dominates and nonsphericity becomes negligible.

We will show now that the friction due to particle production is very strong for natural values of the parameters, so, speaking in the mechanical language, the point particle comes to the origin with vanishingly small angular momentum and almost no charge asymmetry is generated by χ decays. It was believed [7] that particle production is not essential, when χ_1 is large, since the mass of the particles produced is proportional to χ_1. This is not always the case however. The coupling of χ to fermions is of the usual Yukawa type

$$g\chi\,\overline{\psi}_a\psi_b + g^*\chi^*\overline{\psi}_b\psi_a \tag{12}$$

So the effective mass of fermions is of the order of $g\langle \chi_i \rangle$. This value is to be compared with characteristic oscillation frequency (10). Since generically in supersymmetric models $\lambda = O(g^2)$ one may expect almost any result depending upon the exact value of the ratio g^2/λ. If $g^2 > \lambda$ the production rate is exponentially suppressed and in the opposite case it can be large.

To make the statements more explicit let calculate particle production rate by time dependent scalar field χ (t) [13]. Such classical scalar field is equivalent to a time dependent particle mass. We assume that the latter has a constant part m_0 and an oscilling one $m_1(t)$. The explicit analytic formulae are obtained for the particular time dependence

$$m(t) = m_0 + m_1 cos\Omega t \qquad (13)$$

but the formulae are of more general applicability.

If Ω is large in comparison with m_0 and m_1 the production rate can be easily calculated in perturbation theory and coincides with the width of the decay of χ into fermions. If Ω is small perturbation theory is not applicable but fortunately in this case we can use the quasiclassical approximation. If $m_0 > m_1 > \Omega$ an exponential suppression of particle production should take place. As we see below it is indeed the case. If m_0 is small $(m_0 < \Omega)$ and m_1 increases one might naively expect that the production rate also increases since this corresponds to the increasing field amplitude. This is true e.g. for particle production by electromagnetic field. However these naive expectations prove to be wrong for scalar field because the increase of the field amplitude simultaneously leads to the increase of the effective mass of the particles produced. The

net result is a mild suppression of the production rate by the factor $(\Omega/m_1)^{1/2}$.

To do the calculations we start from the classical Lagrangian of a relativistic particle with a variable mass

$$L=m(t)(1-v)^{1/2} \qquad (14)$$

The corresponding Hamiltonian is $H=[p^2+m^2(t)]^{1/2}$. The theory is quantized by the path integral method:

$$G\left(\vec{x}_f, t_f ; \vec{x}_i, t_i\right) = \int \mathscr{D}\vec{p}\,\mathscr{D}\vec{x}\,\exp\left\{i \int_{t_i}^{t_f} dt\,(\vec{p}\,\dot{\vec{x}} - H)\right\}$$

where $G(\vec{x}_f, t_f; \vec{x}_i, t_i)$ is the Green function for particle propagation from the space-time point $(\vec{x}_i, t_i)$ to $(\vec{x}_f, t_f)$.

The integration can be easily done and we get

$$G\left(\vec{x}_f, t_f ; \vec{x}_i, t_i\right) = \int \frac{d^3 p}{(2\pi)^3}\,\exp\left\{i\vec{p}\,(\vec{x}_f - \vec{x}_i) - i \int_{t_i}^{t_f} dt\,\sqrt{\vec{p}^{\,2} + m^2(t)}\,\right\} \qquad (15)$$

It follows from this expression that the pair creation amplitude is

$$A_c\left(\vec{k}_1, \vec{k}_2\right) = (2\pi)^3\,\delta(\vec{k}_1 + \vec{k}_2)\,\exp\left\{-i\int_C dt\,\sqrt{k_1^2 + m^2(t)}\,\right\} \qquad (16)$$

where k_1 and k_2 are the momenta of the created particles. The integration contour in the complex t-plane is chosen in such a way that the energy $(k^2+m^2)^{1/2}$ changes sign along the contour. If $m(t)$ is a periodic function of time the square root has an infinite number of branching points and the total amplitude is the sum of $A_c(16)$ corresponding to all different contours C. The summation results in particular in the δ-function corresponding to the energy conservation. For further details of the method one could see refs.[14,15].

Let first consider the case when m_0 is small. The calcu-

lation is straightforward and we get for the particle produc-
tion rate per unit time and volume:

$$\dot{N} = \frac{2\Omega^2}{\pi} \sum_n \int \frac{d^3p}{(2\pi)^3} \exp\left\{-\frac{4\sqrt{p^2+m_i^2}}{\Omega}[K(\gamma)-E(\gamma)]\right\} \cdot$$
$$\delta\left[n\Omega - \frac{4}{\pi}\sqrt{p^2+m_i^2}\, E\left(\sqrt{1-\gamma^2}\right)\right] \tag{17}$$

Here $\tau = p/(p^2+m_1^2)^{1/2}$ and $K(\tau)$ and $E(\tau)$ are the complete ellip-
tic functions (see, e.g. ref.[15]). The expression becomes very
simple if $m_1 \gg \Omega$:

$$\dot{N} = \frac{\Omega^{5/2} m_i^{3/2}}{2\pi^{7/2} \ell n^{1/2} \frac{m_i}{\Omega}} = \frac{4}{\pi^{5/2}}\sqrt{\frac{\Omega}{m_i \ell n \frac{m_i}{\Omega}}}\; \dot{N}_{PT} \tag{18}$$

where N_{PT} is the production rate in perturbation theory.

When m_o is large the production probability is strongly
suppressed. The analytic calculations are easier in the case
$m^2(t) = m_o^2 + m_1^2 \cos^2\Omega t$. The result is

$$\dot{N} = \sum_n \frac{m_o \Omega^2}{2\pi^3}[m_o(n\Omega - 2m_o)]^{1/2}\exp\left\{-\frac{2m_o}{\Omega}\ell n \frac{16m_o(n\Omega-m_o)}{em_i^2}\right\} \approx$$
$$\approx \Omega^{5/2} m_o^{3/2}\left[\exp\left(\frac{16}{e}\frac{m_o^2}{m_i^2}\right)\right]^{-2m_o/\Omega} \tag{19}$$

These considerations can be also of interest for the Uni-
verse reheating after inflation. For the latter only qualita-
tive estimates have been done [16].

Returning to the baryogenesis I would like to make the
following comment. The sign of baryon asymmetry is determined
stochastically by quantum fluctuations and thus no explicit
C(CP)-violation is necessary. This is a kind of spontaneous
breaking of charge symmetry. The only effect of that is bary-
ogenesis. Particle-antiparticle properties othrewise are the
same.

In this model there are no domain walls which usually appear when a discrete symmetry is broken. The regions with positive baryonic charge are separated from those with negative baryonic charge by baryon-poor space. Their characteristic size can be evaluated as follows. Quantum fluctuations of χ during inflationary stage are

$$\delta\chi = \chi(0) - \chi(L_o) \approx H_I / 2\pi \qquad (20)$$

on the scale $L_o = H_I^{-1}$. This scale expands with time as $L = L_o exp(H_I t)$. The fluctuation amplitude increases in accordance with expression (4) till it reaches limiting value χ_o (6). So the relative fluctuation amplitude becomes of the order of unity when $t_1 = 4\pi^2 \chi_o^2 / H_I^3 = 4\pi^2 const / H_I \sqrt{}$ and correspond ngly the characteristic size of the region with definite baryonic charge to the end of inflation is

$$L_1 = H_I^{-1} exp \left\{ \frac{const}{\sqrt{\lambda}} \right\} \qquad (21)$$

Unfortunately the exact value of the constant under exponent sign is unknown.

The present day size of such a region L_o can be evaluated as follows. During the Friedman stage it expands as $z_R = T_R / T_o$ where $T_o = 2,78$ K is the temperature of the background radiation now and T_R is he reheating temperature after inflation. The latter by the order of magnitude is $T_R^4 = m_{pl}^2 H_I^2$. Hence

$$L_o = Z_R L_1 = \frac{1}{T_o} \sqrt{\frac{m_{pe}}{H_I}} exp \left(\frac{c}{\sqrt{\lambda}} \right) \qquad (22)$$

The observational constraints are satisfied if $(c/\sqrt{\lambda}) > 50$.

The model which is discussed here opens a very interes-

ting possibility of getting a large leptonic asymmetry along with the small baryonic one [13]. This is possible if (B−L) is not conserved and baryonic "friction" due to particle production is much larger than the leptonic one. It is possible to achieve that with a small variation of the coupling constants because the production rate is very sensitive to their values. If this is the case the chemical potential of electronic neutrinos enters into primordial nucleosynthesis calculations as an extra parameter.

It is possible also that the scale of leptonic charge variation is much smaller than that of baryonic charge (22). The helium−4 abundancy in this case should be different in different space points. If this is observed it will be a good confirmation of the model discussed in this talk.

References

1. A.D.Sakharov, Pis'ma ZhETF, 5 (1967) 32.

2. G. t'Hooft, Phys. Rev. Lett. 37 (1976) 8; Phys. Rev. D14 (1976) 3432.

3. V.Kuzmin, V.Rubakov and M.Shaposhnikov, Phys. Lett. B191 (1987) 171;
   M.Shaposhnikov, Nucl. Phys. B. 287 (1987) 757; B299 (1988) 797.

4. S.W.Hawing, Nature 248 (1974) 30;
   Ya.B.Zeldovich, Pis'ma ZhETF, 24 (1976) 29.
   A.D.Dolgov, ZhETF 79 (1980) 337; Phys. Rev. D24 (1981) 1042.

5. A.Cohen and D.Kaplan, Phys. Lett. 199B (1987) 251; Nucl. Phys. B308 (1988) 913.

6. V.A.Kuzmin, Pis'ma ZhETF 13 (1970) 335;

M.Yoshimura, Phys. Rev. Lett. 41 (1978) 281; 42 (1979) 746 (E);

A.Yu.Ignatyev, N.V.Krasnikov,V.A.Kuzmin and A.N.Tavkhelidze, Phys. Lett. 76B (1978) 436.

7. S.Weinberg, Phys. Rev. Lett. 48 (1982) 1303;

M.Khlopov and A.Linde, Phys. Lett. B138 (1984) 265;

J.Ellis, D.Nanopoulos and S.Sarkar, Nucl. Phys. B254 (1985) 175;

M.Kawasaki and K.Sato, Phys. Lett. B189 (1987) 23.

8. I.Affleck and M.Dine, Nucl. Phys. B249 (1985) 361.

9. A.Vilenkin and L.A.Ford, Phys. Rev. D26 (1982) 1231;

A.D.Linde, Phys. Lett. B156 (1982) 335;

A.A.Starobinsky, Phys. Lett. B117 (1982) 175;

10. T.S.Bunch and P.C.Davies, Proc. Astron. Soc. 360 (1978) 117.

11. A.D.Linde, Phys. Lett. 160B (1985) 243;

J.Ellis, E.Enquist, D.V.Nanopoulos, and K.A.Olive, Phys. Lett. 191B (1987) 343.

12. A.D.Dolgov and D.K.Kirilova, JINR preprint P2-89-873 (in Russian). Submitted to Nucl. Phys. B.

13. A.D.Dolgov and D.K.Kirilova. Yadernaya Fizika, 51 (1990) 273 (in Russian); Preprint JINR E2-89-321 (in English).

14. V.S.Popov, ZhETF 61 (1971) 1334; 62 (1972) 1248.

15. I.S.Gradshtein and I.M.Ryzhyk, Tables of integrals, sums, es, and products, Moscow, 1962.

16. L.Abbot, E.Farhi, and M.Wise, Phys. Lett. 117B (1982) 29;

A.Albrecht, P.Steinhardt, M.Turner, and F.Wilczek, Phys. Rev. Lett. 48 (1982) 1437;

A.D.Dolgov and A.D.Linde, Phys. Lett. 116B (1982) 329.

THE COSMOLOGICAL QUARK–HADRON TRANSITION AND ITS PATOLOGIES.

Silvio A. Bonometto

Michele Mariani

Dep. of Physics of the University of Perugia, Via Pascoli, 06112 Perugia (Italy)

I.N.F.N. – Sezione di Perugia

Abstract

When the Universe cooled down to a temperature $\sim 200\,MeV$, strongly interacting matter turned from quark–gluon plasma into hadrons. The behaviour of the Universe during such transition is considered. In particular, we debate the possibility that B can concentrate in small areas. Such isothermal inhomogeneities would later turn into proton inhomogeneities. They may cause primeval nucleosynthesis outputs to be somehow different from those arising in a purely homogeneous framework. The conditions for this to happen are discussed here in some detail. We also discuss the possibility that predictions on Ω_o may change and conclude that it is however quite unlikely that the outcome of the cosmological quark–hadron transition is to reconcile a purely baryonic Universe with $\Omega_o = 1$.

1. Introduction.

The concept of phase transition, formerly relegated in the realm of physics of matter and physical chemistry, has become familiar to particle physicists when the importance of gauge theories was fully appreciated. Gauge vector bosons, associated to conserved quantities through local invariances, would be intrinsically massless. Only in the abelian case could a massive vector field respect gauge invariance (Stückelberg, 1938). But the only abelian gauge field we directly meet is the massless electromagnetic field.

Massless fields seem to imply long range interactions. In the standard model of fundamental interactions, this is avoided requiring that observable particles, in ordinary laboratory experiments, are not the quanta of the fields inscribed in the Lagrangians. Here is where phase transitions become a necessary concept. In the present temperature and density conditions gauge symmetries are broken, but the theory predicts the conditions for symmetry restauration. Then the fields inscribed in the Lagrangian would become directly observable.

Making recourse to the so-called Higgs' mechanism (Higgs, 1964; Kibble, 1967), a phase transition takes place when the degrees of freedom of the *Higgs' scalar field* turn into those of the longitudinal part of a massive vector field. Such mechanism works in the frame of the electroweak unification and is also the basic ingredient for grand-unified approaches. In the former case a transition temperature $T_{ew} \sim$ 100 GeV is coherent with observational parameters. This range of energies is still on the reach of laboratory physics, although in conditions very far from statistical equilibrium. In the latter case transition temperatures $T_{GU} \sim 10^{15}$ GeV are involved and cosmology is required to provide possible tests of the theory.

From a theoretical point of view, the situation is even more extreme for strong interactions. Quantum chromo-dynamics (QCD) is an unbroken gauge theory, although no observable field is set in its Lagrangian. Baryons and mesons (hadrons), which are the protagonists of laboratory strong interactions and the main ingredient of nuclear matter, are to be interpreted as composite states. Hadron hard cores can be easily interpreted as proper volumes and this is typical of non-elementary objects. In spite of that no laboratory experiment can separate hadrons into their components.

This is due to a basic assumption of QCD, the so-called *confinement*. Spinor and vector fields of the QCD Lagrangian have quanta which ought to remain *confined* inside hadrons. This is related to the $SU(3)$ quantum numbers which define *colour*. Colour charge is the source of forces which, unless rapidly *saturated*, would diverge with distance. The *saturation* distance itself is approximately the hard-core size

$s_{hc} \sim 10^{-13} cm$. Average distances of this order – already not much below those in nuclear matter – may become typical in two different context.

A former case are neutron (N) star cores, that we shall not discuss here. Cosmology is the other case where hadrons are expected to form only after an earlier epoch when strongly interacting matter (SIM) was in the form of quark–gluon plasma. This case differs from the former one being characterized by a very high ratio $\eta = S/B \sim 10^8$ between entropy (S) and baryon number (B).

It is not difficult to evaluate the order of magnitude of the temperature T_c for which SIM turns form quark–gluon plasma into hadrons. At $T > m_\pi$ (m_π: π meson mass), the π number–density (neglecting π mass, volume and mutual interactions) is $n_\pi = \zeta(3)/\pi^2 \, 3 \, T^3 \simeq 1.16 \lambda_\pi^3 (T/m_\pi)^3$. This shows that distances between quarks belonging to the same or to different nearby π's become similar already for T slightly above m_π. Therefore, at a temperature above $\sim 200 \ MeV$, we expect that the natural form for SIM is a quark–gluon plasma.

A further characteristic, which differenciates the quark–gluon plasma from the hadron gas, is the behaviour in respect to chiral symmetry. Quark spinors have a small *electroweak* mass. Baryon mass, instead, is originated by strong interactions. Therefore, the passage from quark–gluon plasma into hadron gas is marked both by the onset of confinement and by chiral symmetry breakdown.

A reasonable possibility exists that an experimental insight into its physics of Q–H transition becomes available in a foreseable future. This is different from the situation holding for grand–unification and even from the electroweak transition itself. In fact the Q–H transition can be hopefully reproduced in laboratory, in conditions quite near to statistical equilibrium, at least for short time intervals.

In this paper we wish to outline some patterns along which astrophysical data and QCD physics can find a mutual interaction. As a consequence of quark–hadron transition primeval nucleosynthesis may take place in the presence of significant inhomogeneities of the proton (P) distribution. It is however important to stress two points: i) It seems quite unlikely that P inhomogeneities can lead to a coherent framework consistent with $\Omega_B = 1$. ii) In spite of that, possible modification of the homogeneous nucleosynthesis results lead to a possible relaxation of the limit $\Omega_B \leq 0.1$. Further evaluations of the limit on Ω_B, suitably varying inhomogeneity data, can be important.

Some of the early approaches to the cosmological quark–hadron transition, while rich of innovative ideas, were essentially aimed to find quick results. But, also in this field, precise results need detailed work. This is true also for the question of Ω_B limits, which can only be discussed on the light of a full comprehension of

what went on when the Universe cooled down to a few hundreds MeV. In this paper we shall try to debate the present understanding of this problem, adding some new contributions and trying to focus their weight in the general outline of current ideas.

2. Nature of the quark–hadron transition.

At a temperature $T_c \sim 200\ MeV$ the quark–gluon plasma turns into hadron gas. An important point is whether this is really a phase transition. In principle there might exist a T interval during which both hadrons and quark–gluons are present. Quite in the same way as during hydrogen (re)combination, when the ionization rate x gradually shifts from $x \simeq 1$ at $T \gtrsim 5000\ K$ down to $x \simeq 10^{-5}$ at $T < 2500 K$.

(a) – Order of the transition.

However, even though the passage from quark–gluons into hadrons is a phase transition, it is essential to know its order. Only a first order phase transition (for which $p_h(T_c) = p_q(T_c)$ while $\frac{\partial p_h}{\partial T}(T_c) \neq \frac{\partial p_q}{\partial T}(T_c)$; p_h and p_q are the pressures of the two phases) can be expected to cause really significant cosmological effects. Different techniques have been applied to try to find out the order of the transition.

Renormalization group techniques are not fully reliable in this context, being however based on a perturbative approach whose results are to be applied in a regime where the effective coupling constant $\alpha_s(T) \sim 1$. The pressure of the quark–gluon plasma calculated in this way becomes negative below a suitable $\bar{T}$ (see Kalashnikov and Klimov 1979 and Kapusta 1979). This indication of a first order phase transition is to be taken, as already stressed, with reserve.

A better insight into the physics of the transition is obtainable from Montecarlo lattice computations. This approach has been used in various steps. Former computations, neglecting fermions or with non–dynamical fermions, initially using $SU(2)$ as colour group, then the full $SU(3)$, indicated a first order phase transition related to confinement. Relevant technical difficulties were to be overcome to treat finite mass spinors. When they were solved, no evidence of a first order phase transition connected with confinement was found if dynamical fermions have masses so low as those of light quarks. However it become soon evident that a first order phase transition might be present because of the chiral symmetry breakdown.

We should add that more recent analyses, obtained by the APE collaboration (Bacilieri et al. 1988, 1989) completely reopened the debate on the order of the transition, claiming that no phase transition existed even in the presence of pure gluons. These outputs were criticised by Ukawa (1989) and Karsch (1989). According to private communications, it now seems that the most recent APE collab-

oration results agree with initial findings. In this context it is however reasonable to assume that a first order phase transition did take place.

Montecarlo outputs express T_c in terms of the so-called *string tension* Λ which is known only with an approximation $\sim 30\%$. As a consequence, however, T_c itself is not known better than the heuristic arguments given in the previous section and based on eq. (1.2) could provide it.

(b) – Thermodynamical models of phases.

It is clear that, in the vicinity of the transition neither hadrons nor quark–gluons behave as ideal gases. We can model the expected behaviour of hadrons taking into account their intrinsical volume. At $T \gtrsim m_\pi^{-1}$, hadron surfaces are almost in contact. Inter–hadron forces can then be quite significant. However, even if these forces are neglected and we take account of the intrinsical volumes only (Karsch and Satz 1980, Bonometto 1983), number density, energy density and pressure are to be modified in respect to a perfect gases.

The quark–gluon phase can hardly be analysed starting from a full QCD treatment. Models like the MIT bag model and generalizations can then be a valid operational approach.

For both phases it is convenient to define

$$\phi_{h,q}(T) = p_{h,q}(T) T^{-4}, \qquad E_{h,q}(T) = \rho_{h,q}(T) T^{-4} \qquad (2.1)$$

Similar quantities can be defined for the ν component as well. Then

$$E_q(T_q) = \pi^2/30\{g_r + g_{(q)}[1 + (n/3 - 1)(\bar{T}/T_q)^n]\} \qquad (2.2)$$

$$\phi_q(T_q) = \pi^2/90\{g_r + g_{(q)}[1 - (\bar{T}/T_q)^n]\} \qquad (2.3)$$

$$E_h(T_h) = 3\phi_h(T_h) = \pi^2/30[g_r + 3f_{rid}] \qquad (2.4)$$

$$E_\nu(T_\nu) = 3\phi_\nu(T_\nu) = \pi^2/30 g_\nu \qquad (2.5)$$

where $g = \mathcal{N}_b + \frac{7}{8}\mathcal{N}_f$ ($\mathcal{N}_{b(f)}$: number of boson (fermion) spin states), the index $_r$ refers to leptons and photons, the index $_{(q)}$ refers to quarks and gluons, f_{rid} accounts from the difference between an ideal gas and a real pion gas, while n and $\bar{T}$ allow to choose different self-consistent quark–gluon models. In general $\bar{T} = T_c[1 - \frac{3f_{rid}}{g_{(q)}}]^{(1/n)}$. For the bag model $n = 4$ and the bag constant $B = (\pi^2/90)g_{(q)}\bar{T}^4$. As is known the bag model yields an $E(T)$ decreasing with T. MonteCarlo lattice outputs do not seem to agree with this behaviour. $E(T)$ is increasing for $n < 3$. In what follows we shall therefore consider both the cases $n = 4$ and $n = 2.7$.

(c) – The onset of the transition.

Owing to the large value of η, T_c can be determined equating the pressures of the two phases. In $Fig.1$ typical Φ_q and Φ_h behaviours are given. They are coherent with the requirement

$$\rho_q(T_c) > \rho_h(T_c). \tag{2.6}$$
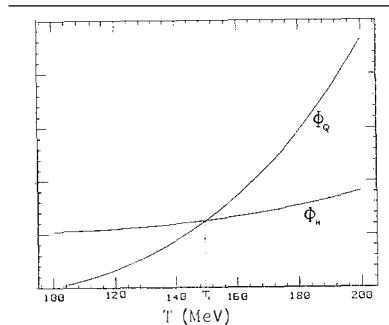
$Fig.\,1$ – Typical behaviours of pressures around T_c.

When the Universe cools down to T_c, its expansion can continue isothermally, while the fraction $y = V_q/V_{tot}$ (V_{tot} is a large volume in order that homogeneity conditions hold on its scale; V_q is the part of V_{tot} occupied by the Q–phase) gradually increases from 1 to 0. The most convenient volume sharing, from a thermodynamical point of view, has been calculated by Bonometto and Sakalleriadou (1984; see also Bonometto and Pantano 1984, Bonometto and Matarrese 1983, Lodenquai and Dixit 1983, Bonometto and Masiero 1986), who also gave the $a(t)$ behaviour during the transition and showed that, in these conditions, entropy is conserved.

However, there are at least two kinds of reasons why the transition cannot take place in thermodynamical equilibrium. A certain amount of supercooling, in respect of T_c, is required to drive energy from the quark volume to the hadron volume. As can be seen from $Fig.1$, requiring $p_h = p_q$ at $T < T_c$, yields $T_h < T_q$. The amount of this *physiological* supercooling depends on a number of conditions and will be discussed in the next section.

Most phase transitions, e.g. between liquid and gas, start with the nucleation of bubbles of the new phase inside the pre-existing phase. In this case one should expect hadron bubbles to nucleate in the quark–gluon continuum as soon as T_c is

attained. However bubble formation takes place when there is a positive surface energy density at the boundary between the two phases. Should the creation of boundaries imply a net absorption of energy, dendritic structures may form. This is still an open possibility, although unlikely. In this case, most expected cosmological effects of the transition would not take place.

In the case of bubble formation, the number of bubbles nucleated at $T < T_c$ is

$$n_{nuc} = d_{nuc}^{-3} = exp[\mathcal{K}/(T_c - T)]. \tag{2.7}$$

$\mathcal{K}$ is a constant. Once a bubble nucleates, the latent heath warms up the surrounding quark region. T increases until the volume warmed ($V_w = d_w^3$) becomes too large and the cooling effect of the universal expansion takes over again. Meanwhile bubble nucleation in V_w is suppressed. When the cooling process arrives to a temperature T_{sc}, for which $d_w \simeq d_{nuc}$, bubble nucleation ends.

In order to determine T_{sc} (and the corresponding average distance d_1 between nearby nucleation sites) one needs to follow in detail the process of heath feeding in the volume V_w, while the initial bubble expansion takes place. This stage of the transition has been studied in numerical details by Miller and Pantano (1989). This results, however, still depend on the surface tension σ.

Kajantee and Karkkainen (1988, see also Kajantee *et al.* 1989 and Frei and Patkós 1989) tried to evaluate σ on the basis of lattice computations. They find

$$\sigma = \sigma_o T_c^3 \tag{2.8}$$

with $0.1 \lesssim \sigma_o \lesssim 10$. (In spite of such wide computational uncertainties, it seems however that this excludes dendritic structure formation.)

Still in wait of final conclusions on T_{sc} and d_1, various possibilities will be considered here. It is however unlikely that the supercooling $T_c - T_{sc}$ is greater than the *physiological* supercooling described in the next section. This larger *patological* supercooling prevents the transition from taking place – in its initial stages – along the pattern corresponding to a minimal entropy increase. Other possible *patologies* could arise if elementary processes exist whose rates do not match the conditions for thermodynamical equilibrium. An example could be the rates of processes transferring B–number from the Q–phase to the H–phase. The relative abundances of B are prescribed by chemical equilibrium requirements, but if the rates of elementary reactions are inadequate, chemical equilibrium is not reached.

While the transition can take place even in the presence of pure volume processes (Bonometto and Pantano, 1987), the presence of surface processes is essential to

allow B–transfer. This is a reason to study which relative contribution surface and volume processes can acquire. Then, among the former ones, the weight of B–transferring processes shall be considered.

(d) – Surface processes during the transition.

Once a substancial volume of the Universe is occupied by hadronic gas, the energy transfer between phases takes place through both surface and volume processes.

Most processes are bound to occur in a thin layer about the phase boundaries (surface processes). Some of them (type A) will be scatters between two particles each belonging to a different phase. If $T_q \neq T_h$, type A processes produce a net energy transfer. Rearrangements inside each phase follow such transfer.

Furthermore we expect processes like

$$q + \bar{q} \; \rightarrow \; \pi, \qquad q + \bar{q} \; \rightarrow \; 2\,\gamma, \qquad q + \bar{q} \; \rightarrow \; e^+ + e^-$$

(type B) to realise an actual matter transfer. Most SIM in the Q–phase will turn into leptons and photons. In fact, at $T < T_c$, $g_{qg} = 37$, while, at $T < T_c$, $g_h \simeq 3$. Meanwhile the statistical weight of leptons and photons has no change.

For each type A or type B process, however, the collisions per unit time and volume is $\nu_{rs} = < \sigma_{rs} v_{rs} > n_r\, n_s$. Here $_{r,s}$ indicate particle species, σ_{rs} and v_{rs} are cross section and impact velocity between r-th and the s-th particles. The averaging is over possible kinematics and output products. Finally, $n_{r,s}$ are particle number densities. The average energy transfer per collision is $\xi_{rs}(T_q - T_h)$ with $0 < \xi_{rs} \lesssim 1$. The average energy transfer per unit time and volume, due to surface processes, is then $f \frac{\Sigma \lambda_c}{V_t} \left(\frac{\rho_{h_c}}{T_c} \right)^2 \frac{T_q - T_h}{T_c^2}$ where $\lambda_c = T_c^{-1} \sim 10^{-13} cm$ and

$$f \; = \; \frac{s}{\lambda_c} \sum_{r,s} \xi_{rs} f_r f_s < \frac{\sigma_{rs}}{\lambda_c^2} v_{rs} > . \qquad (2.9)$$

Here $f_r = n_r T_c / \rho_{h_c}$, s is the depth of the layer where elementary processes occur, $\frac{\Sigma}{V_t}$ is the average inter–phase surface per unit volume (see the end of the next section). Eq. (2.9) sets the connection between the hydrodynamical rates and the elementary processes rates *plus* the description of each particle species. Unfortunately, for most quantities in it, only estimates are available . This suggests to treat f as a parameter in a suitable range of values.

(e) – Volume processes.

Around T_c the mean–free–path of ν is $\lambda_\nu \simeq 0.5(m_\pi/T)\,cm$ and this is likely to be compatible with bubble sizes. Scatterings of ν's inside each phase (volume processes) lead to energy sharing between ν and collided particle. In average this will give place to an energy transfer similar to type A surface processes.

The smallness of the numbers

$$N_{q,h} = \frac{4}{3}f_{q,h}G_w^2\rho_\nu(\rho_{q,h} + p_{q,h}) \tag{2.10}$$

($G_w^2 = 1.3\,10^{-22}\ Mev^{-4}$ is the weak interaction constant; $f_{q,h} \sim 1$, see next section) of collisions of ν's with quarks and hadrons (per unit time and volume) is compensated by the fact that the whole volume V_{tot} is to be considered.

3. Basic dynamical conditions.

Considering in detail the effects of the processes discussed in section 2-c, we can work out the t dependence of scale factor a, volume sharing $y = V_q/V_{tot}$ (V_q: volume occupied by the Q–phase on a given large volume V_{tot}), temperatures T_q, T_h, T_ν. (Clearly T_ν has a meaning over scales $>> \lambda_\nu$.) This can be obtained integrating a system of differential equations that we shall now report.

(a) – Dynamical equations.

Space averaging the equation $T_{k;i}^i = 0$ (T_k^i is the stress–energy tensor) in a synchronous gauge, we obtain

$$\sum_{i=h,q,\nu} [(\rho_i + p_i)\dot{V}_i + \dot{\rho}_i V_i] = 0 \tag{3.1}$$

Here q, h and ν refer to the Q–phase, the H–phase and ν's; ρ_i are energy densities, p_i pressures and V_i volumes taken by the i-th component. Neglecting B and considering both surface and volume processes between the phases, we have

$$\dot{\rho}_q + (\rho_q + p_q)\frac{\dot{V}_q}{V_q} = -(T_q - T_\nu)N_q - f\frac{\Sigma\lambda_c}{V_t}\left(\frac{\rho_{h_c}}{T_c}\right)^2\frac{T_q - T_h}{T_c^2} \tag{3.2}$$

$$\dot{\rho}_h + (\rho_h + p_h)\frac{\dot{V}_h}{V_h} = -(T_h - T_\nu)N_h + f\frac{\Sigma\lambda_c}{V_t}\left(\frac{\rho_{h_c}}{T_c}\right)^2\frac{T_q - T_h}{T_c^2} \tag{3.3}$$

On the r.h.s. of these equations, the former term accounts for volume effects, the latter one for surface effects. In the volume terms $N_{q,h}$ is given by eq. (2.10). The factors $f_{q,h} \simeq 1$, in (2.10), take account of the different relations between density

and number–density in the two phases and of the ratio between temperature difference and average energy transfer per collision. The surface terms [the same in (3.2) and (3.3)] account for energy transfer in the two opposite directions. Together with (3.2) and (3.3) the Friedman equation

$$\left(\frac{\dot{a}}{a}\right)^2 = \left(\frac{8\pi}{3}\right) G[\rho_h + \rho_\nu + (\rho_q - \rho_h)y] \tag{3.4}$$

holds, while the pressure equality

$$p_h = p_q \tag{3.5}$$

will be used to relate T_h and T_q. This holds at rest in respect to phase boundaries. Boundaries, however, move with velocity $v \simeq (d_1/t) \simeq c(d_1/s_h)$. Correcting for this effect, one easily sees that the pressure balance can be shifted by an amount $\Delta p \simeq (\rho_q - \rho_h)(d_b/s_h)^2$. This is to be compared with the typical shifts of p_q and p_h from the value p_c holding at T_c, that we shall compute. If the condition $d_1 \lesssim 10^2 \, cm$ is satisfied, Δp turns out to be neglegible.

(b) – Analytical solution of the dynamical problem.

Bonometto and Pantano (1987) found an analytical solution to the above system, considering volume terms only. This was based on an expansion in respect to

$$\alpha = \frac{3}{t_1} \frac{1}{G_w^2 T_c^5} = 3 \, 10^{-5} \left(\frac{m_\pi}{T_c}\right)^5 (8.3 \, 10^{-6} s/t_1) \tag{3.6}$$

after which only terms $\boldsymbol{\bullet}(\alpha)$ are taken into account. Here t_1 is the time when the transition begins. The procedure can be extended, taking into account surface terms as well. The time dependence of the scale factor a and of the volume ratio y is still the same here as in Bonometto and Matarrese 1983, where no temperature shift in respect to T_c was considered. Changes for $a(t)$ and $y(t)$ are $\boldsymbol{\bullet}(\alpha^2)$. These solutions read

$$\left(\frac{a}{a_1}\right)^3 = U_c \sin^2 \left[(6\pi\phi_H)^{\frac{1}{2}} \frac{T_c^2}{m_{pl}}(t - t_1) + \arcsin(U_c^{-\frac{1}{2}})\right]$$

$$y = \left(\frac{a_1}{a}\right)^3 (1 + A_c) - A_c \tag{3.7}$$

(the index c refers to transition temperature conditions; $a_1 = a(t_1)$, $U = 1 + \frac{E_Q}{\phi_Q}$, $A_c = \frac{E_H + \phi_H}{(E_{(q)} - E_{(h)})_c}$; the indeces $H(Q)$ are used to indicate the whole contents in both phases; e.g., $E_Q = E_q + (\pi^2/30)g_\nu = E_{(q)} + (\pi^2/30)(g_r + g_\nu)$).

Let us now define the fractional temperature shifts

$$q = \frac{T_q - T_c}{T_c}, \qquad h = \frac{T_h - T_c}{T_c}, \qquad \nu = \frac{T_\nu - T_c}{T_c} \qquad (3.8)$$

which are $\mathcal{O}(\alpha)$. Their expressions are fairly complicated and read

$$q = \frac{4\phi_{q_c}}{3\phi_{q_c} - E_{q_c}} \left[\frac{\pi(y) + \mu(y)}{1 + \tau(y,a) + \omega(y,a)} \right] \qquad (3.9)$$

$$h = \frac{E_{q_c} + \phi_{q_c}}{3\phi_{q_c} - E_{q_c}} \left[\frac{\pi(y) + \mu(y)}{1 + \tau(y,a) + \omega(y,a)} \right] \qquad (3.10)$$

$$\nu = \mu(y) + \frac{\pi(y) + \mu(y)}{1 + \tau(y,a) + \omega(y,a)} \left[\frac{E_{q_c} + \phi_{q_c}}{3\phi_{q_c} - E_{q_c}} - \omega(y,a) \right] \qquad (3.11)$$

Here

$$\pi(y) = A(y)(E_H + \phi_H)/(f_q y), \qquad \mu(y) = A(y)(E_Q + \phi_Q)_c/[f_h(1-y)] \qquad (3.12)$$

with

$$A(y) = \alpha\Gamma \frac{\bar{E}^{\frac{1}{2}}(y)}{(E_{(q)} - E_{(h)})_c(E_\nu + \phi_\nu)} \qquad (3.13)$$

$(\bar{E}(y) = E_h + E_\nu + (E_{(q)} - E_{(h)})_c y)$ are related to ν convection. The quantities

$$\tau(y,a) = S(y,a)[f_q y (E_q + \phi_q)_c]^{-1}, \qquad \omega(y,a) = S(y,a)[f_h(1-y)(E_h + \phi_h)]^{-1} \qquad (3.14)$$

with

$$S(y,a) = \alpha \frac{f t_1}{3} \frac{\Sigma}{V_t} \frac{E_{h_c}^2}{E_\nu + \phi_\nu}, \qquad (3.15)$$

instead, are related to processes taking place through the surface Σ limiting the two phases in the volume V_t. The expressions (3.8), taking account of the definitions (3.9) – (3.15), are taken from Mariani and Bonometto (1990). These equations solve the differential problem. The next subsection is devoted to a discussion of the dependence of $\tau(y,a)$ and $\omega(y,a)$ on $\frac{\Sigma}{V_t}$. Then we shall plot the time dependence of T_i, as obtained from above relations, in a number of different cases, and shall briefly comment about such outputs.

(c) – Evolution of inter–phase surfaces during the transition.

The transition takes place through the following steps: i) After nucleation, bubbles expand and eventually collide. Then two bubbles turn into a single one. ii) When $\sim 40\%$ of the volume is occupied by H–phase, surface rearrangements make it the physically connected phase. (Let us call the steps i) and ii) *first stage* (1S) of the

68

transition.) iii) Q–bubbles shrink down [*second stage*: (2S)]. All through the above steps, *i.e.* bubbles are assumed to be spherical.

It is theoretically possible that d_1 is so large that the stage i) never really occurs and $\sim 40\%$ of V_t gets occupied by H–phase before bubbles actually meet. In this case, the final distance between nearby quark–bubble centers is $\sim d_1$. Otherwise $d_2 \gg d_1$.

It is

$$\frac{\Sigma}{V_t} = (36\pi)^{\frac{1}{3}}\mathcal{N}_h^{\frac{1}{3}}(1-y)^{\frac{2}{3}} \quad (1S), \qquad \frac{\Sigma}{V_t} = (36\pi)^{\frac{1}{3}}\mathcal{N}_q^{\frac{1}{3}}y^{\frac{2}{3}} \quad (2S) \qquad (3.16)$$

Here $\mathcal{N}_{h,q}$ are the numbers per unit volume of hadron or quark bubbles and, in (2S), $\mathcal{N}_q = d_2^{-3}(a_2/a)^3$. A similar evolution takes place early in (1S), before bubble collision, when $\mathcal{N}_h = d_1^{-3}(a_1/a)^3$. Later, when bubbles begin to touch, $\mathcal{N}_q$ is controlled by the distribution of distances (d) about d_1. In this paper such distribution is assumed to be essentially Gaussian. More details on the way $\mathcal{N}_q$ varies are given in Mariani and Bonometto (1990).

(d) – Evolution of temperatures during the transition.

According to the expressions (3.9), (3.10), (3.11), we can plot the behaviours of T_q, T_h, T_ν during the whole transition. Such behaviour depends on the equation of state used for the phases (see eq. 2.5). For the H–phase this reduces to fixing $f_{rid} = 0.3$. For the Q–phase, characterized by rapidly variable conditions, we take the equations corresponding to $n = 4$ (bag-model) and $n = 2.7$. In Figs. 2–5 the r.h.s. plot corresponds to $n = 4$, the l.h.s. plot to $n = 2.7$.
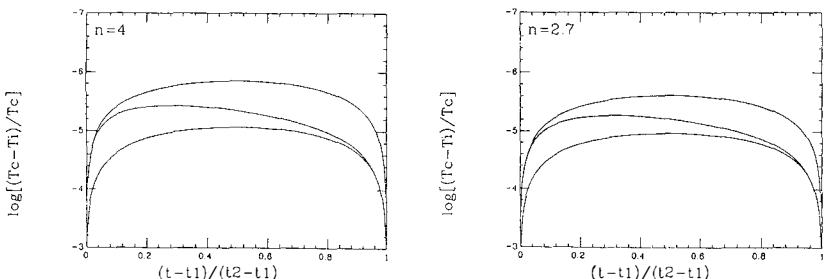
Fig. 2 – Temperature behaviours for $f = 0$, *i.e.* in the absence of surface processes.
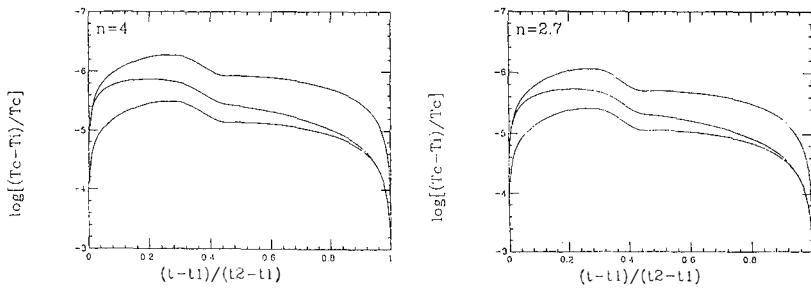
$Fig.\,3$ – Temperature behaviours for $f/d_1 = 10^{-7}$, $f/d_2 = 10^{-8}$
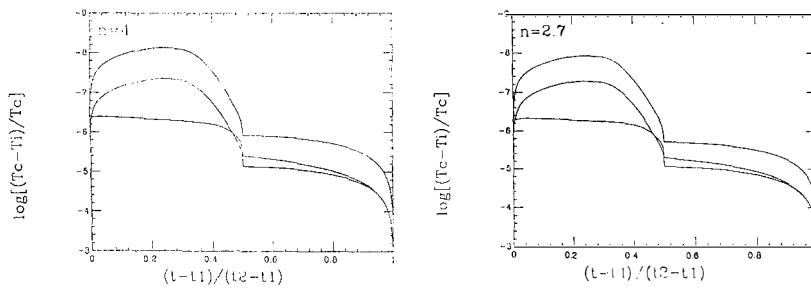
$Fig.\,4$ – Temperature behaviours for $f/d_1 = 10^{-5}$, $f/d_2 = 10^{-8}$
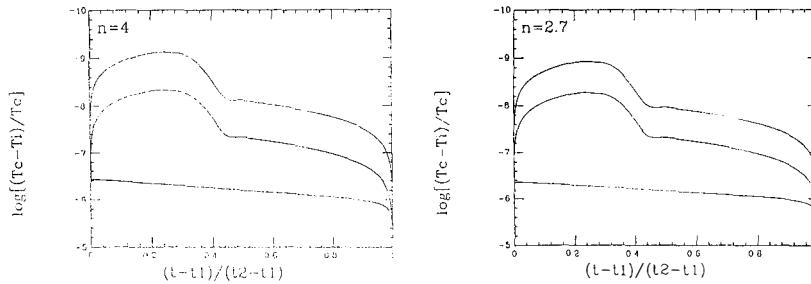
$Fig.\,5$ – Temperature behaviours for $f/d_1 = 10^{-4}$, $f/d_2 = 10^{-5}$

The parameters f/d_1 and f/d_2 mix together bubble dynamics (through the values of d_1 and d_2) and surface process rates (see eq. 2.9).

The ν temperature can either lay between T_h and T_q or be below both of them. Only in the former case ν contribute to the energy transfer between phases.

Among other conclusions which can be drown from the above plots, we see which range of parameters allow to mantain the temperatures of the phases very near

to T_c until the transition is over. There have been suggestions, in the literature, that quark–gluon plasma areas might undercool significantly, in the late transition stages. The above results seem to exclude this perspective.

4. B flows during and after the transition.

(a) – B accumulation during the transition.

That B can tend to remain in the Q–phase during the transition was first outlined by Witten (1984). This led him to introduce the so–called *quark nuggets*, about which we shall not discuss any longer. Witten arguments were however reelaborated by Bonometto *et al.* (1985) and Applegate and Hogan (1985).

A first reason leading to B concentration is related to the fact that baryons have a mass $>> T_c$, while light quarks are lighter than T_c. This point can be made quantitative if precise models for the two phases are introduced and, in particular, spurious effects due to the continuum spectrum

$$\rho(m,B) = \sum_{k=1}^{K} g_k \delta_{B,B(k)}\, \delta_{m,m_k} + c\,\theta(m-m_o)\, T_c^{2.5}\, m^{-3.5}\, e^{m/T_c - 7\pi^2 B^2 T_c/30m} \quad (4.1)$$

(m_k: discrete hadron masses; $m_o \simeq 1200\,MeV$: starting point for the continuum spectrum; $4 \lesssim c \lesssim 40$) are cured considering a finite volume (e.g., $V_o = (4\pi/3)m_{pi}^{-3}$) model for hadrons (Karsch and Satz, 1980; Bonometto, 1983) leading to the partition function

$$\mathcal{Z}_{T,\mu,V} = \frac{1}{8\pi^3} \sum_N \frac{1}{N!} \prod_{i=1}^{N} \sum_{B_i=-1}^{+1} \int d^3 q_i dm_i \rho(m_i, B_i) e^{[-(E_i+B_i)/T]}(V - 4NV_o)^N \quad (4.2)$$

(Here the sum over N is carried up to a value $\simeq V/4V_o$; μ is the chemical potential associated to B and $E_i^2 = q_i^2 + m_i^2$.) From $\mathcal{Z}_{T,\mu,V}$ all quantities concerning the H–phase can be worked out. In Bonometto *et al.* (1985) this is used to work out the hadron free–energy, which is then compared with the quark-gluon value. At the equilibrium temperature the chemical equilibrium requirement provides the value for the ratio $\epsilon = < B >_H / < B >_Q$. This turns out to depend on T_c itself and on the other detail of the model; its typical values range around 10^{-2}. Further details can be found in Bonometto *et al.* (1985).

Witten (1984) considered also the possibility that surface processes can be highly suppressed during (2S), as $\Sigma_q << \Sigma_h$. But, according to the results of the previous section, this does not occur. The result of this hypothetical suppression would have been those of leaving ν's as the only carriers of energy off the plasma remnants. In this way B would keep trapped in the late shrinking bubbles. Also in the presence

of surface process this might still happen if the rate of B–carrying processes is much lower than the rate of other surface processes. A possible reason for such diminished rate is the need to provide somehow a 3–body reaction to create a baryon out of quarks.

This pattern can be somehow eased by the possible existence, at least as a virtual state, of a di-quark (DQ) system. Then the chains of reactions, causing B–flow through the inter–phase surface, might be the following ones:

$$C_1: \quad Q + Q \to DQ; \quad Q + DQ \to B, \qquad C_2: \quad Q + Q \to DQ; \quad DQ \to B + \bar{D}Q$$

The knowledge of DQ is however still limited and makes hard to give even an estimate of the rates of the reactions here above. They might be smaller, by some order of magnitude, than ordinary surface process rates. This kind of patology, though preserving the behaviour of temperature trends and volume sharing deduced in sec. 4, would drive out from the chemical equilibrium condition requiring $\epsilon \simeq 10^{-2}$.

Another kind of pathology, causing similar effects, would be the accumulation of B in the proximity of the inter–phase surface, even after it has passed from the Q–phase to the H–phase. In this case, chemical equilibrium would be established between the interior of Q–bubbles and surronding B–enriched areas. A detailed evaluation of this effect is again problematic, for it involves estimates of superimposed hydrodynamical flows and B diffusion.

On the light of previous consideration, it is however licit to conclude, that – at the end of the Q–H transition, we can expect a situation of B inhomogeneity. Peaks of B will have formed in the regions were late plasma bubbles died out. The value of ϵ characterizing such areas is likely to exceed 10^{-2}, even by a few orders of magnitude.

(b) – B outflow after the transition.

If the transition causes the formation of B peaks, they are later unstable against N diffusion. Until $T > 1\,MeV$, each baryon can be considered P or N for 50% of its life. In a time t, baryons the diffuse over a distance $s \simeq (Dt)^{\frac{1}{2}}$ with

$$D = (90/4\pi^2)(g_r T^3 \sigma_N)^{-1} \qquad (4.3)$$

(g_r is the statistical weight of radiative matter), while the N elecromagnetic cross–section $\sigma_N \simeq \frac{17}{81}\sigma_T \left(\frac{m_e}{m_q}\right)^2$ (the value of m_q to be taken conserns u and d flavours only).

If, at the end of the transition (started at a time t_o), B peaks had a distance d_2, the ratio between the size of B peaks and their distance, at any $T << T_c$, reads then

$$\frac{s}{d} \simeq \frac{t_o}{10^{-5}\,s}\left(\frac{1\,MeV}{T}\right)^{\frac{3}{2}}\left(\frac{1\,mb}{\sigma_N}\right)^{\frac{1}{2}}\frac{1\,cm}{d_2} \qquad (4.4)$$

This relation shows that, if d_2 could exceed the cm scale (or, at least, for those late Q–bubbles whose distance from nearby objects exceeded such scale), a B accumulation is however preserved down to the T when $P - N$ transforming reaction freeze down. But, even for values $s/d \sim 10^{-2}$, proton inhomogeneities can still be caused. The detailed picture will depend on a number of geometrical parameter as well on the actual value to be taken for the *initial* ϵ.

At $T < 1\,MeV$, N leak out freely, but P are definitely trapped and nucleosynthesis can be expected to take place in the presence of significant P distribution inhomogeneities. Unfortunately, at this stage, the theory is still unable to give firm details on the features and on the amplitude of such inhomogeneities. These quantities will be necessarily treated as working parameters.

5. Nucleosynthesis and quark–hadron transition

Homegeneous big–bang nucleosynthesis has been deeply studied and there can be little doubt that its results are a cornerstone of cosmology. Light element observed abundances agree with its prediction in a range of 9 orders of magnitude. Its prediction of the number of elementary particle families came well before the same result could be achieved in LEP Z^o decay experiments. Using data concerning H^2, He^3 and Li^7, the baryonic density parameter Ω_B is strongly constrained.

A recent discussion of the observational situation and its comparison with theoretical prediction based on recent rates of nuclear reactions, is contained in Kurki–Suonio *et al.* (1989). For homogeneous nucleosynthesis, the full consistency interval is claimed to be

$$3 \leq \eta_{10} \leq 4 \qquad (5.1)$$

where $\eta_1 0 = (n_B/n_\gamma) \times 10^{10}$ (n_B, n_γ: baryon and photon number densities in the present epoch Universe). In turn $\eta_{10} \simeq 1.5\,10^2\Omega_B h^2$ (h: present value of the Hubble parameter in units of $100\,Km\,s^{-1}Mpc^{-1}$), and this yields $\Omega_B \lesssim 0.1$.

Current limits on Ω_B, obtained along this pattern, take also into account the possible range of values of h. Lower and upper limits on Ω_B appear to differ by a factor ~ 8. If observational constraints are indicated by (5.1) feelings can arise that further modest variations of experimental rates in nuclear reactions can dangerously narrow, or even close, the residual window. The same danger is related to possible variations in the observational data on light nuclide abundances. The

residual window, however, can be not so narrow, if a number of actual uncertainties are duly taken into account (Krauss and Romanelli, 1990).

The importance of evaluating the effects of a nucleosynthesis taking place in the presence of P inhomogeneities, are also to be seen under this light. However, even though the *homogeneity* window is expected to keep open, it aquires a critical significance to verify which upper limit on Ω_B can be altogether inferred from observed abundancies.

In homogeneous nucleosynthesis, the so–called *Deuterium bottleneck* becomes accessible when the ratio between N and P numbers is $n : p \simeq 1 : 7$. The presence of P density peaks leads in turn to vaste areas of proton deficit, where $n : p$ easily exeeds unity. The opening of *Deuterium bottleneck* has the *normal* consequence of fixing all available neutrons into H^2 nuclides. But, should areas whith $n : p > 1$ exist, once all protons are set into H^2 nuclides, a neutron residual will still exist. To find P partners they have to flow back into P peak areas, as protons are virtually inamovable. While this flow–back occurs, N decay will go on. Protons produced at this stage will soon meet a N partner. However, apart these events, an extra–time

$$\Delta t \simeq \frac{d^2}{D} \tag{5.2}$$

will be needed in order that each N meets its P partner [D is given in (4.3)].

The effect is potentially similar to a slower expansion of the Universe and makes a large Ω model to look like a smaller Ω one. Large d_2 tend to favour this effect. If we treat d as a free parameter, however, we must be very careful not to exceed obvious limits, above which the expected amount of ^4He would be altered. Keeping below this limit areas where only H^2 exists and other areas with significand H^2 deficit are however generated.

A general picture of the consequences of this situation can be fully reached only on numerical bases.

According to Applegate *et al.* (1987) and to Alcock *et al.* (1987), in these conditions the observed abundancies of He^4, He^3 and H^2 could become compatible with $\Omega_B = 1$. The theoretical amount of Li^7, instead, exceeds observed abundancies. The existence of hypothetical processes able to destroy such nuclide was then advocated. Audouze, Delburgo–Salvador and Salati (1989) and Reeves (1989) argued against such possibility and rather tried to infer from the observed Li^7 abundancies limits on the possible outputs of the quark–hadron transition.

Furthermore Malaney and Fowler (1988) stressed that these computatiuons treated the high P density regions and the low P density regions separately. An accurate

evaluation of nucleosynthesis must take into account all possible neutron flowings and, in particular, must be very accurate on the study of diffusion processes.

Taking them into account, one could hopefully be able to eliminate the excess Lithium originated via Be^7 decay after that the latter nuclide had been originated in high density regions. This led to studying detailed diffusion models (Kurki–Suonio *et al.*, 1989; Teresawa and Sato, 1988) showing that, beside eliminating Be^7, the back neutron diffusion affects other nuclides as well, and to build a detailed dynamical code (Kurki–Suonio and Matzner, 1989) explicitly accounting for multizone forward and backward diffusion.

Kurki–Suonio *et al.* (1989), making use of such dynamical code, explored a vaste range of possible conditions ensuing the quark–hadron transition. The parameters accounting for such possible conditions were so defined:

R: ratio between P densities inside and outside P peaks.

d: average distance between nearby P peak centers.

f_v: average volume fraction taken by P peaks.

Together with these parameters, also various $\eta_1 0$ parameter values were span.

Attempts were also performed to account for the possible geometries of P peaks, but no attempt was made to account for the predictable variations in P densities through the so–called P peaks. Sharp peak boundaries were also used and this approximation cannot be considered completely safe *a priori*, while it might be advisable to consider at least steps in the variation of R.

We shall now report the observational constraints taken into account by Kurki–Suonio *et al.* (1989).

$$0.224 \leq \frac{\rho_{He^4}}{\rho_{tot}} \leq 0.254$$

$$\frac{n_{H^2}}{n_H} \leq 10^{-5} \qquad \frac{n_{H^2+He^3}}{n_H} \leq 10^{-4}$$

$$\frac{n_{Li^7}}{n_H} \leq 2\,10^{-10} \;\; (popII) \qquad \frac{n_{Li^7}}{n_H} \leq 2\,10^{-9} \;\; (popI)$$

The range of parameters they considered is $R \sim 100$, $1\ m < l < 10^3 m$, $0.016 < f_v < 0.25$, $1 < \eta_1 0 < 10^2$.

On the basis of such choice of parameters they find an overall limit

$$\eta_{10} \lesssim 7 \tag{5.3}$$

this rises the current limit on Ω_B to ~ 0.2.

It must be however considered that this increase takes place at the expences of considering $d \sim 100\,m$ (at $T \sim 100\ MeV$). It is hard to reconcile such high value of d with current extimates of surface tension.

In our opinion it might be more significant to look for alterations of homogeneous nucleosynthesis considering smaller values of d but higher values of R. The value of R considered by Kurki–Suonio *et al.* (1989) is consistent with the B concentration ratio between quark and hadron areas, which can be worked out during the quark–hadron transition from the assumption of chemical equilibrium (Bonometto *et al.*, 1985). As outlined in the previous sections, however, the intervention of patologies at this stage is far from unlikely, and $R \simeq \epsilon^{-1}$.

Altogether it seems useful that different areas in the parameter space are soon explored. Teresawa and Sato (1988) claimed that values $R \sim 10^3, 10^4$ are consistent with $\Omega_B = 1$. A similar claim by Mathews *et al.* (1989) seems, instead, to have been withdrown.

Higher values of R may hopefully lead to alterations of the homogeneous picture even for smaller d values. It seems unlikely that this may reconcile nucleosynthesis and $\Omega_B = 1$, but a shift of the upper boundary on Ω_B by a factor ~ 3 does not seem out of reach.

It is known that $\Omega_B \neq 1$ is required by a number of different astrophysical evidences. *E.g.*, it is definitely difficult to reconcile $\Omega_B = 1$ and current limits on small scale fluctuations of the CBR. On the contrary, the CDM model, very successful to account for a large deal of cosmological evidences, is not in agreement with some very large scale data, although its disagreement is just above the limit acceptable on the basis of observational uncertainties. There have been proposals trying to improve the results of CDM models adding a substantial amount (clearly exceeding 10%) of baryonic component to them.

However, even apart the results of specific cosmological models, it is important to connect limits on the density of the Universe and properties of SIM. In our opinion, future work in this field is likely to set even narrower connections between such astrophysicals and laboratory data.

76

References

Alcock C., Fuller G.M. and Mathews G.J., 1987, Ap. J. **320**, 439.

Applegate J.H., Hogan C.J. and Scherrer R.J., 1987, Phys. Rev. D **35**, 1151.

Audouze J., Delbourgo–Salvador P., Salati P., 1988 – "Dark Matter" Proc. of the XXIIIrd Recontre de Moriond, eds. J. Audouze and J. Tran Thanh Van (Editions Frontiers), 227.

Bacilieri P. *et al.*, 1988, Phys. Rev. Lett. **61**, 1545.

Bacilieri P. *et al.*, 1989, Nucl. Phys. B **318**, 553.

Bonometto S.A., Marchetti P.A. and Matarrese S., 1985, Phys. Lett. A **157**, 216.

Bonometto S.A. and Masiero A., 1986, Riv. del Nuovo Cimento vol. **9** n. **5**.

Bonometto S.A. and Matarrese S., 1983, Phys. Lett. B **133**, 77.

Bonometto S.A. , 1983, Nuovo Cim. **74 A**, 325.

Bonometto S.A. and Pantano O., 1984, Astron. Astrophys. **130**, L43.

Bonometto S.A. and Pantano O., 1987, Astron. Astrophys. **176**, L9.

Bonometto S.A. and Sakellariadou M., 1984, Ap. J. **282**, 370.

Frei Z. and Patkós A., 1989, Phys. Lett. B **222**, 469.

Higgs P.W., 1964, Phys. Lett. **12**, 132; Phys. Rev. Lett. **13**, 508.

Kajantie K. and Kärkkäinen L., 1988, Phys. Lett. B **214**, 595.

Kajantie K., Kärkkäinen L. and Rummukainen K., 1989, preprint HU–TFT–89–29.

Kalashnikov O.K. and Klimov V.V., 1979, Phys. Lett. B **88**, 328.

Kapusta J.I., 1979, Nucl. Phys. A **148**, 461.

Karsch F. and Staz H., 1980, Phys. Rev. D **12**, 1168.

Karsch F., 1989, preprint CERN–TH–5498/89 (to appear in Quark–Gluon Plasma", ed. R.C. Hwa (World Scientific).

Kibble T.W.B., 1967, Phys. Rev, **155**, 1554.

Krauss L.M. and Romanelli P., 1989, preprint YCTP–P1–89.

Kurki–Suonio H. and Matzner R.A., 1989, Phys. Rev. D **39**, 1046.

Kurki–Suonio H., Matzner R.A., Olive K.A. and Schramm D.N., 1989, Fermilab–Pub 89/252–A; Ap. J. in press.

Lodenquai J. and Dixit V., 1984, Phys. Lett. B **147**, 273.

Mariani M. and Bonometto S.A., 1990, Ap.J. (submitted)

Mattews G.J., Fuller G.M., Alcok C.R., and Kajino T., 1988, "Dark Matter" Proc. of the XXIIIrd Recontre de Moriond, eds. J. Audouze and J. Tran Thanh Van (Editions Frontiers), 319.

Melaney R.A. and Fowler W.A., 1988, Ap. J. **333**, 14.

Miller J.C. and Pantano O., 1987, Phis. Rev. D **40**, 1789.

Reeves H., 1988, "Dark Matter" Proc. of the XXIIIrd Recontre de Moriond, eds. J. Audouze and J. Tran Thanh Van (Editions Frontiers), 287.

Stüskelberg E.C.G., 1938, Hel. Phys. Act. **11**, 299.

Teresawa N. and Sato K., 1988, University of Tokyo Preprints: UTAP77, UTAP79, UTAP83.

Ukawa A., 1989, preprint UTHEP–199

LATE-TIME PHASE TRANSITIONS AND LARGE SCALE STRUCTURE

Christopher T. Hill

David N. Schramm

The University of Chicago,
5640 S. Ellis Avenue, Chicago, IL 60637
and
NASA/Fermilab Astrophysics Center,
Fermi National Accelerator Laboratory
Box 500, Batavia, IL 60510-0500

Lawrence M. Widrow

Harvard University and Center for Astrophysics
60 Garden Street, Cambridge, MA 02138

ABSTRACT

The Universe may have undergone a vacuum phase transition subsequent to the decoupling of the microwave background radiation with matter. Under certain circumstances, "soft" cosmic topological defects can form, such as domain walls, which can lead to the formation of large scale structure. Since these structures form after decoupling, the constraint imposed by the observationally small anisotropy of the microwave background radiation, $\delta T/T \lesssim 10^5$, is weakened. A soft-defect scenario is a novel alternative to both cosmic string and inflation-produced quantum fluctuations as the origin of structure in the Universe.

The problem of generating struture (galaxies, clusters, voids, stars, people, etc.) in a universe that appears very homogeneous and isotropic on the largest distance scales and earliest times is, perhaps, the central problem in cosmology today. The difficulty faced by any proposed mechanism owes to the constraint on the initial "lumpiness" of the universe from the observed smoothness of the $3°K$ microwave background radiation. This radiation, a relic of the big bang, last interacted with matter at a redshift of about $z \sim 10^3$, and shows a surprising uniformity in temperature with fluctuations bounded by as little as $\delta T/T \leq 2 \times 10^{-5}$ on some angular scales.[1] The lumpiness of the universe at this time is bounded as well, since pre-existing density fluctuations would have differentially red- and blue-shifted the background radiation, giving rise to observable temperature fluctuations. It is well known that small initial density fluctuations, $\delta\rho/\rho \ll 1$, can only grow linearly with the cosmological expansion, $[(\delta\rho/\rho)_{today} \propto (1+z)(\delta\rho/\rho)_{redshift\,z}]$, due to gravitation. How, then, did the enormous density contrasts observed today grow in the period of $\sim 10^3$ redshifts, given that the current $\delta T/T$ limits seem to imply $(\delta\rho/\rho)_{z\sim10^3} \lesssim 10^{-4}$ at decoupling?[2]

Once the density fluctuations are of order unity, they grow very rapidly and can produce the contrasts seen, for example, in comparing the density of a star to the average density of matter in the universe, $\rho_{star}/\rho_{ambient} \sim 10^{28}$. If linear growth started at decoupling, then $(\delta\rho/\rho)_{today}$ is $\lesssim 0.1$, implying that we would never have attained nonlinear growth. On the other hand, if there is non-baryonic dark matter (*e.g.*, axions), growth could start when $\rho_{matter} \sim \rho_{radiation}$ at $z \sim 10^4$, which might produce the start of non-linear growth at $z \sim 1$. This may seem to be marginal, yet one might argue that the spectrum of initial density fluctuations contains a small probability of having few large fluctuations that can ultimately lead to the formation of the observed structure. However, from the existence of quasars and galaxies at large redshiifts we see that well-formed structure already exists at $z \sim 4$. Furthermore, there appears to be a large-scale coherent streaming motion of galaxies[3] on a scale of $R \sim 50$ *to* $100Mpc$, suggesting the existence of exceedingly massive objects. Such objects are very difficult to form in most models given the present limits on $\delta T/T$. If further observations ultimately yield

a tightening of the limit on $\delta T/T$, the situation will become even more constrained.

Previous proposals for the formation of large-scale structure have relied upon generating density fluctuations at a very early cosmological epoch (*e.g.* the Grand Unified [GUT] epoch when $kT \sim 10^{14} GeV$) which survive to serve as seeds at the galaxy formation epoch at $kT \sim 10^{-2} eV$. These include quantum mechanical Gaussian fluctuations produced during inflation[4] and topological defects such as cosmic strings.[5] In some scenarios these seeds gravitationally accrete large quantities of non-baryonic dark matter, whereas in others they explode and push the baryons about.[6] We will not go into a detailed commentary on each of these models, noting only that the aforementioned combination of observations has been difficult (but maybe not imposible) for any existing model to satisfy.

The purpose of this article is to review and update a completely novel, if not radical, proposal[7] in which energy density fluctuations are generated *after* decoupling.[8] This implies *a priori* a minimal imprinting of the induced structure upon the microwave background radiation, *i.e.*, a relatively small induced $\delta T/T$ for any given produced structure. The fluctuations here are associated with "soft" topological structure, typically in the form of domain walls (though this is not the only possibility) having small internal energy densitites. The domain walls are kink-soliton, topologically stable solutions to the equations of motion for some very weakly interacting scalar field, ϕ. Random vacuum fluctuations with large spatial scales are also a possibility. In the models considered, the mass of the ϕ particles m_ϕ is so small that the thickness of the kink, given by the Compton wavelength, $\hbar/m_\phi c$, is a cosmological distance scale. The original motivation for expecting such low-mass particles and a late-time phase transition came from a study of possible pseudo-Goldstone bosons that naturally arise in a variety of GUT settings. Pseudo-Goldstone bosons, such as massless familons[9], occur when the pattern of masses of the observed fermions is associated with a spontaneously broken, continuous, global (ungauged), symmetry. With further small explicit breakings of these symmetries, familons acquire minuscule masses, as, for example, in the "schizon" models[10]. These masses are typically of order $m_\phi \sim m_f^2/f_\phi$, where $f_\phi \sim 10^{14} GeV$ to $10^{16} GeV$ is

a generic grand unification scale, and m_f the mass of the associated family of fermions.

In ref. [7] a specific particle-physics based model was considered which postulated the existence of such schizons in association with the neutrinos. This tied the central density of a domain wall to the mass of a neutrino, m_ν, and the thickness of the wall to both m_ν and a grand-unification scale f_ϕ. For $f_\phi \sim 10^{15} GeV$ and m_ν in the range 1.0 to 0.01 eV, the ratio of wall density to the ambient density of matter will then become greater than unity at a redshift of $z \sim 100$. For the above parameters, the thickness of the wall will be in a range of 10 to 10^5 parsecs. The phase transition sketched out in ref. [7] is fundamentally no different than those invoked in inflationary schemes. The key idea is that new physics is introduced here involving phenomena of extremely low energies: much of what we say is generic to any late-time phase transition. Indeed, it is interesting to explore further the possible connections with particle physics that harbor such phenomena.

In the remainder of this discussion we will simply treat the domain walls as phenomenological objects having characteristic thicknesses, δ, and mass per unit area, σ (for the original model of ref. [7], $\delta \sim f_\phi/m_\nu^2$ and $\sigma \sim m_\nu^4 \delta \sim m_\nu^2 f_\phi$). The walls form during a phase transition at a redshift z_0. In addition, we define R as the typical correlation scale for structures in the domain wall network. At the time of the phase transition, the walls are randomly distributed and typically contiguous or intersecting, with the average spacing between walls, r_0, of order $r_0 \sim \delta$ to $r_0 \sim H^{-1}$ depending upon the model. As the Universe expands and cools, the system relaxes, and individual walls become well-defined kink-soliton configurations. The spacing between walls grows and becomes $r_o(1 + z_0)/(1 + z)$ at a redshift $z < z_0$. In addition, there will be slow recombination of structures as well as other evolutionary effects. The domain wall network is expected to contain both closed surface walls (dubbed "vacuum bags") and infinite walls.

Ultimately, one expects infinite cosmic domain walls to become flat on the scale of the horizon. If the walls have a large central energy density, then they give rise to unacceptably large distortions in the microwave background and Hubble flow; this is

the "usual" cosmic domain wall disaster and was first noted by Zel'dovich *et al.*[11] In
the case of soft walls, the central energy densities are very low and the large domain
walls suggest an intriguing mechanism that may account for the large-scale streaming
motion in a natural way,[12] as we shall describe below. It is furthermore expected that
small local structures, *e.g.*, vacuum bags, which are spherical bubbles whose walls are
the kink-soliton, will form and become, ultimately, the nucleation sites for galaxies,
etc. Several groups[13,14,15] are now actively analyzing the details of this scenario, *e.g.*,
behavior of vacuum bags, the evolution of domain wall networks, and the distribution
of observable structures expected in the model.

The evolution of the walls and vacuum bags is quite complicated as there are poten-
tially a very large number of processes that can come into play. The stress-energy of a
wall consists of a surface-energy density and a surface-tension *of equal magnitude.* This
surface tension causes vacuum bags to collapse and small-scale irregularities on infinite
walls to oscillate. In either case, the walls lose energy via ϕ-particle and gravitational
radiation. In addition, if there is a background density of some other particles that
interact with ϕ (*e.g.*, neutrinos), then it will exert a force that will tend to damp any
motion of a wall relative to the cosmic rest frame.

Press, Ryden and Spergel[13] have completed a preliminary analysis of the evolution
of a domain wall network. They find that the network quickly becomes dominated by
infinite walls that are flat on scales of order the horizon and that small local structures
quickly disappear. We note, however, that their numerical code may lack the resolution
necessary to track the ultimate fate of the vacuum bags. From the point of view of
structure formation, the fate of vacuum bags is the most interesting question. It is
important to note, moreover, that the evolution depends crucially upon the underlying
effective theory of the field which produces the domain wall kinks. In a theory with a
potential of the form:

$$V(\phi) = \lambda(v^2 - \phi^2)^2 \tag{1}$$

i.e., a "double-well" potential, the walls tend to "intercommute," undergoing a rear-

rangement upon interacting, and dissipating energy in the form of free ϕ, and some gravitational, radiation. Vacuum bags in these models shrink down to blobs that eventually disperse into free particles.[13,14] Left behind are very large walls which stretch across the entire universe. Naively, if such structures are related to galaxy formation scenarios (thereby requiring $\rho_{wall}/\rho \gtrsim 10^{-4}$), then they would lead to a large $\delta T/T$ in conflict with observations. We view this as a less interesting scheme However, if the model is of the theoretically favored[7,10,14] "sine-Gordon" type, with:

$$V(\phi) = m^4 \sin(\phi/f) \tag{2}$$

as always occurs in pseudo-Goldstone boson theories, then there are remarkable stability constraints on the kink-solitons: two opposing, flat kinks in a collision are transparent and will pass through one another. This implies that domain walls tend not to intercommute, unless they have large curvature, and leads moreover to a striking behavior for vacuum bags, recently demonstrated by Widrow.[14] A spherical vacuum bag will collapse and undergo a "bounce"; in this process a small fraction of the energy is lost to radiation. Typically the vacuum bag reexpands and continues to recollapse for many iterations until finally only a dissipative blob, or even a black hole, remains. The stability of the vacuum bag can no doubt be enhanced by endowing it with anisotropies, inhomogenieties, angular momentum, *etc.*

We should further point out that the flat infinite domain wall can be, in principle, avoided by making all domain walls unstable, as occurs in various incarnations of the models discussed above. Consider, for example, the superposition of sine-Gordon potentials with multiple non-degenerate minima:

$$V(\phi) = m_1^4 \sin(n_1\phi/f) + m_2^4 \sin(n_2\phi/f) \tag{3}$$

where $n_1 < n_2$ are integers. If n_2/n_1 is noninteger, we expect domain walls to occur, but one side will be a region of false vacuum having higher vacuum energy than the true vacuum on the other side of the wall. Regions of false vacuum shrink due to vacuum pressure, and all walls eventually disappear. However, the density fluctuations due to

the walls may live long enough to drive structure formation. This alternative has not been explored in detail.

The bounce behavior of vacuum bags is important, since a vacuum bag with $\rho_{bag}/\rho >$ 1 (irrespective of whether it contains true or false vacuum) persisting for a Hubble time can drive the nonlinear accretion of the surrounding matter. Moreover, a highly anisotropic bag would be expected to develop self-interesting points of large curvature which are expected to lead to "fission" into smaller vacuum bags, *etc.* Thus, a parenting process for the formation of local clusters of galaxies can be envisioned here in which the matter ultimately accretes onto the remaining small vacuum bags in a neighborhood defined by the large parent structure. Simulations of the evolution of complex vacuum bags have not yet been carried out, but work is in progress. In the numerical simulations of both Press *et al.*,[13] and Widrow,[14] only surface tension and ϕ radiation are taken into account. The motivation for this is both simplicity and the belief that graviational radiation and particle-wall interactions are negligible for the models of interest. Still, these other processes should be investigated in more detail.

Flat cosmic domain walls, owing to the presence of internal pressure as well as energy density, actually gravitationally *repel* matter.[16] On the contrary, spherical vacuum bags have net vanishing pressure and positive mass as seen from outside at distances greater than the radii of the bubbles, and will attract matter. Accreting vacuum bags lead to subsequent evolution of the conventional matter with collapse times that are much more rapid than standard linear growth. Furthermore, the energy within a vacuum bag at the end of its history could mimic the effects of dark matter. In the central core of a collapsing bag the energy density is of order $\sigma(R_0^2/\delta^3)$ where R_0 is the initial radius of the bag. Here, all of the initial surface energy in the bag has been localized into a region of radius $\sim \delta$. One can even form a black hole if the Schwarzschild criterion is satisfied.[14]

The flat, repulsive walls may actually help explain certain puzzling cosmological observations, as envisioned by Stebbins and Turner.[12] As mentioned above, a remarkable coherent streaming motion of all local galaxies within a region $\sim 100 Mpc$ has

been observed[3] which is extremely difficult to explain in conventional scenarios. It has been proposed that a "Great Attractor,"[17] a super-super-cluster of order $10^{17} M_\odot$, in the general direction of Hydra-Centaurus may be required. However, in ref. [12] it was suggested that this could arise from the great domain wall stretching across our present Hubble volume. Moreover, the arguments of Stebbins and Turner can be used to place a limit on the fraction of critical density in the ϕ field today, Ω_ϕ, due to the induced large-scale velocities. From the present data[3] on $R \sim 40 Mpc$, we know that $\Omega_\phi (1 + z_0)\delta\rho/\rho \lesssim 0.2$, where z_0 is the redshift of the phase transition and $\delta\rho/\rho$ is the density variation in the ϕ field. Thus, for $\delta\rho/\rho \sim 1$ we have $\Omega_\phi \lesssim 0.2/(1 + z_o)$. This constraint sets bounds on the evolution of the ϕ field structures.

The fluctuations in the microwave background for a late-time model will usually be dominated by the effect of the propagation of the background through the gravitational field set up by the moving wall network. One finds:

$$\frac{\delta T}{T} \sim \frac{\gamma v}{c} G\sigma R \qquad (4)$$

where v is the velocity of the structure, γ is the relativisitic factor, and G is Newton's constant. For vacuum bags, $R \sim \delta$ and $\gamma \sim 1$, so the approximate result of reference [7] is obtained. Note that $\delta T/T$ increases with R to the maximum structures produced. Therefore, $\delta T/T$ rises with angular size θ until encompassing the maximum scale R, and then remains flat for larger θ. For sufficiently small σ, the above expression for $\delta T/T$ is supplanted by the Rees-Sciama[18] effect from the observed structure, but such values of σ are unlikely to generate structure. In general, the walls are massive enough that $v/c \leq 1$, and $\gamma \sim 1$. If evolution leads to a few dominant walls, then $R \sim R_H$ and the large $\delta T/T$ problem arises as described by Press et $al.$[13] or possibly a situation similar to the one considered by Stebbins and Turner[12] might arise.

Multiple minima or non-degenerate minima with decaying walls can lead to $R \leq R_H$. If there exists an observable structure of size R, then late-time walls will tend to produce the minimum $\delta T/T$ consistent with such a structure. However, if evolution implies even larger structures as in ref. [13], one can encounter limiting constraints. Thus, scenarios

which lead to $R \lesssim 100 Mpc$ are preferred. Note that if vacuum bags are produced, then the structure they seed can be independent of the large wall structure, in which case σ can be reduced sufficiently even to enable $G\sigma R_H$ to be consistent with $\delta T/T$. For the original model of ref. [7] with R arbitrary (instead of $R \sim \delta$), we obtain:

$$\frac{\delta T}{T} \sim \frac{m_\nu^2}{M_{Planck}} f\phi R \sim 2 \times 10^{-6} \left(\frac{m_\nu}{10^{-2}ev}\right)^2 \left(\frac{f_\phi}{10^{15}GeV}\right) \left(\frac{R}{10Mpc}\right) \tag{5}$$

which could easily accomodate structures of $R \sim 100 Mpc$ for reasonable parameter values.

Tests for the model vary with the specific details. The mechanism of ref. [7] requires a generic pseudo-Goldstone boson which will be hard to detect directly, but its brethren associated with charged leptons or quarks produce potentially observable new phenomena, e.g., new compostion-dependent pseudo-Gravitational forces, as detailed in ref. [10]. The observation of such effects and non-zero neutrino masses would be compelling circumstantial evidence for possible cosmological effects proposed here.

Obviously, much work remains to be done to examine the details of this class of models. In particular, the astrophysics of the detailed large-scale structure that is generated by such late-time fluctuations is only sketched here; and full hydrodynamic calculatins will have to be carried out. Furthermore, detailed particle physics models will have to be developed to see if all the preferred properties really exist in a fully consistent model. Eventually we would hope to make detailed quantitative predictions about the model vis-à-vis large-scale structure. However, the present large-scale structure observatrions are still quite qualitative. Quantitative statistical measures have yet to describe definitively the apparent structure in a reproducible manner. Anecdotally, voids, filaments, sheets, bubbles or sponges appear,[19,20] depending on the analyses used and on the rapporteur. Conceivably, cosmic membranes could make any or all of these structures depending on how they evolve. Hopefully, specific quantitative predictions will be made before the observational data converge. Our purpose here is to alert readers to the fact that an alternative to the standard galaxy formation scenarios may exist. The physics it relies upon is not any more exotic than the GUT physics that the

standard scenarios utilize. At low energy scales, the model might even be testable in the laboratory. In any case, it may be the only model that can survive possible eventual limits of order $\delta T/T \lesssim 10^{-6}$.

Acknowledgements

References

1. J. M. Uson and D.T. Wilkinson *Astrophys. J.* **277**, L1 (1984).

2. N. Vittorio and J. Silk *Astrophys. J.* **285**, L39 (1984) and *Phys. Rev. Lett.* **54**, 2269 (1985)

   J. R. Bond and G. Efstathiou *Astrophys. J.* **285**, L45 (1984).

3. A. Dressler, S.M. Faber, D. Burstein, R.L. Davies, D. Lynden-Bell, R.J. Turkevich, and G. Wegner *Astrophys. J.* **313**, L37 (1987

4. S. W. Hawking *Phys. Lett.* **B115**, 295.

   A. A. Starobinsky *Phys, Lett.* **B117**, 175 (1982).

   A. H. Guth and S. Y. Pi *Phys. Rev. Lett.* **49**, 1110 (1982).

   J. M. Bardeen, P.S. Steinhardt, and M.S. Turner *Phys. Rev.* **D28**, 679 (1983).

5. Ya. B. Zel'dovich it Mon. Not. R. Astr. Soc. **192**, 663 (1980).

   A. Vilenkin *Phys. Rev. Lett.* **46**, 1169 (1981), 1496 (E).

6. J. P. Ostriker, C. thompson, and E. Witten *Phys. Lett.* **B180**, 231 (1986).

7. C. T. Hill, D. N. Schramm, J. Fry *Comm. on Nucl. and Part. Phys.* **19**, 25 (1989).

8. The generation of structure during a phase transition *after* decoupling was considered by I. Wasserman *Phys. Rev. Lett.* **57**, 2234 (1986). However, the formation of, and cosmological effects due to, topological defects was not considered.

9. F. Wilczek *Phys. Rev. Lett.* **49**, 1549 (1982).

10. C.T. Hill and G.G. Ross *Phys. Lett.* **203**, 125 (1988); *Nucl. Phys.* **B311**, 253 (1988).

11. Ya. B. Zel'dovich, I. Yu. Kobzarev, and L. B. Okun *Zh. Eksp. teor. Fiz.* **67**, 3 (1974); *Sov. Phys. JETP* **40**, 1 (1975).

12. M. S. Turner and A. Stebbins *Astrophys. J. Lett.* **339**, L13 (1989).

13. W.H. Press, B.S. Ryden, D.N. Spergel *Astrophys. J.* **347**, 590 (1989).

14. L. M. Widrow *Phys. Rev.* **D40**, 1002 (1989).

15. H. Hodges *Phys. Rev.* **D39**, 3557 (1989).

    L. Kawano *Phys. Rev.* **D41**, 1013 (1990).

16. P. Sikivie and J. Ipser *Phys. Rev.* **D30**, 712 (1984).

    L. M. Widrow *Phys. Rev.* **D39**, 3571 (1989).

17. D. Lynden-Bell, S. M. Faber, D. Burstein, R. L. Davies, A. Dressler, R. J. Terlevich, and G. Wegner *Astrophys. J.* **326**, 19 (1988).

18. M. J. Rees and D. Sciama *Nature* **217**, 511 (1968).

19. V.de Lapparent, M. J. Geller, and J. P. Huchra *Astrophys. J.* **302**, L1 (1986).

    J. R. Gott III, A. L. Melott, and M. Dickinson *Astrophys. J.* **306**, 341 (1986).

    M. P. Haynes and R. Giovanelli *Astrophys. J.* **306**, L55 (1986).

20. T. Broadhurst, R. Ellis, D. Koo, and A. Szalay *Nature* **343**, 726 (1990).

NEW RESULTS FROM BIG BANG NUCLEOSYNTHESIS

Keith A. Olive
School of Physics and Astronomy
University of Minnesota
Minneapolis, MN 55455

ABSTRACT

Recent results from big bang nucleosynthesis are presented. Changes in 1) some of the nuclear cross-sections; 2) measurements of the neutron half-life and 3) recent measurements of ^{4}He and ^{7}Li abundances, continue to support the success of the standard model. Consistency for all the light elements is achieved for $N_\nu = 3$ when the baryon to photon ratio, $\eta = 3\text{-}4 \times 10^{-10}$.

The standard big bang model for nucleosynthesis is well known[1-3] and I will not review it here. Instead, I will concentrate on the effects of new data: nuclear cross-sections, the neutron mean-life, and ^{4}He and ^{7}Li observations. The standard model remains the simplest theory, which can account for the observed abundances of all the light elements, D, ^{3}He, ^{4}He and ^{7}Li.

Historically, the prime element to test the success of a cosmological model is ^{4}He. In the standard model, the ^{4}He abundance Y_p, depends on basically three "parameters": 1) the baryon-to-photon ratio η; 2) the neutron mean-life τ_n; and 3) the number of light particle degrees of freedom, commonly taken to be the number of neutrino flavors. As we will see, all of these parameters including η are reasonably well determined. The number of (light) neutrino flavors is now limited to three by LEP.[4] Although LEP does not exclude particles without Z^o interactions, any such superweakly interacting particles can be constrained by nucleosynthesis.[5] The neutron mean-life is now quite well known primarily due to recent measurements of Mampe et al.[6] of $\tau_n = 887.6 \pm 3$ sec. Finally, recent data and analyses of ^{4}He and ^{7}Li when combined with previous data from D and ^{3}He restrict η to values $(3-4) \times 10^{-10}$. As such, the fraction of closure density in baryons is restricted to $\Omega_B = 0.01 - 0.1$.

I will begin with a brief discussion of changes in the nuclear reaction rates. The new compilation of nuclear rates by Caughlan and Fowler[7] show four rates which differ significantly from those used by Kawano, Schramm and Steigman (KSS).[8] They are: D(d,n)^{3}He (up by 30%); D(d,p)T (up by 30%); ^{4}He(t,γ)^{7}Li (down by 20%); and ^{7}Li(p,α)^{4}He (up by 10%). Relative to the older work by Yang et al[1], KSS found increases in ^{7}Li primarily at low values of η. This was due to an increase in the ^{7}Li production rate ^{3}He(α,γ)^{7}Li and a decrease in the destruction rate ^{7}Li(p,α)^{4}He. The ^{7}Be rates which affect ^{7}Li at high η were essentially unchanged. The new data lowers the ^{7}Li abundance at low η while increasing it at higher η. The latter effect is due to the increase in the D(d,n)^{3}He rate which leads to more ^{7}Be. In addition the ^{7}Be(n,p)^{7}Li has decreased which also leads to more ^{7}Li at high η. (^{7}Li is more fragile than ^{7}Be.) The effects of uncertainties in the nuclear reaction rates was recently examined by Krauss and Romanelli.[9]

The neutron mean-life has recently been determined to very high accuracy by Mampe et al.[6], using a glass storage container coated by Fomblin oil. This container has a very low probability for leakage

$(2-3 \times 10^{-5}$/bounce at room temperature) for ultra cold neutrons with kinetic energies $E_n \leq 10^{-7}$ eV. The measured mean lifetime in this experiment is $\tau_n = 887.6 \pm 3$ sec.

Prior to this measurement, the world average value of τ_n has been steadily decreasing from $\tau_n = 918 \pm 14$ from the measurement of Christensen et al.[10] in 1972 to its present value of

$$\tau_n = 889.6 \pm 2.9 \text{ sec.}$$

This average is based on the measurements of τ_n listed in the table below.

918	± 14	Christensen	72
903	± 13	Kosvintsev	86
891	± 9	Spivak	88
876	± 21	Last	88
887.6 ±	3	Mampe	89
877	± 10	Paul	89
878	± 30	Kossakowski	89
893.6 ± 5.3		Byrne	90

(References for these measurements can be found in the review of Particle Properties[11] with the exception of the most recent measurement by Byrne et al.[12].

The results of the big bang nucleosynthesis calculations for the new rates and new value of τ_n are found in Figures 1-3 for ^{4}He, D, ^{3}He and ^{7}Li as a function of η.[13,14] The dependence of Y_p on η, $\tau_{1/2}$ and N_ν can be seen in Fig. 1. In Figures 2 and 3 the nucleosynthesis predictions for D, ^{3}He and ^{7}Li (by number) are shown as a function of η. The helium abundance as a function of η, τ_n and N_ν is well fit by

$$Y_p \simeq 0.0228 + 0.010 \ln \eta_{10} + 0.012 (N_\nu - 3)$$

$$+ .185 \left(\frac{\tau_n - 889.6}{889.6} \right) \tag{2}$$

where $\eta_{10} = \eta/10^{-10}$.

I turn now to the observations of ^{4}He and ^{7}Li. For a discussion on the D and ^{3}He isotopes see refs. 1,2,14. Recently, Pagel[15] has put together a large data sample of 33 extragalactic HII regions. In
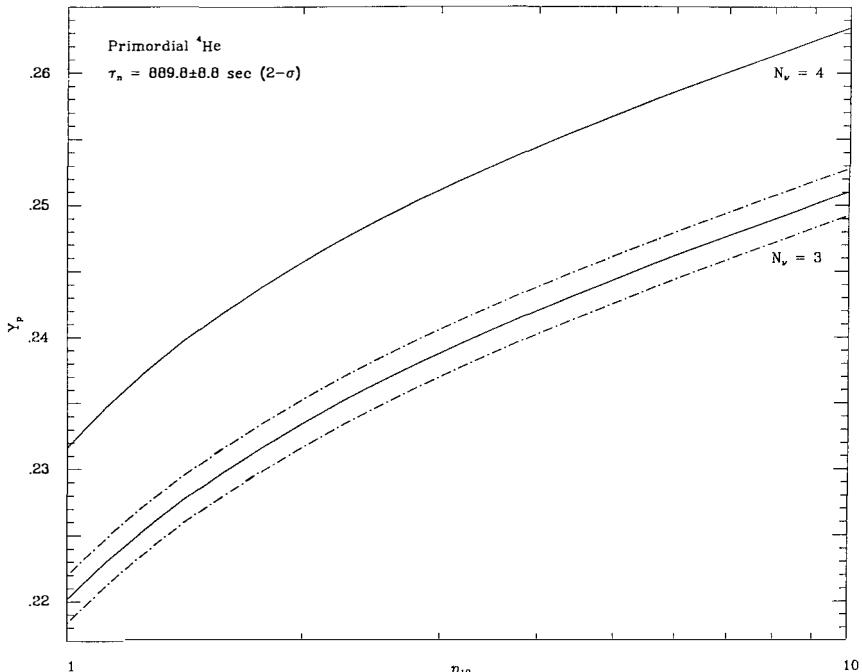
Figure 1

preparing this sample, Pagel has reanalyzed that data (starting from the
spectral line strengths) using a common set of atomic data to deduce
abundances. He has also included a correction based on collisional
excitation[16] which lowers the ^{4}He abundance. Three of these objects have
been updated in ref. 17. The data set used in the figures below have
included the updated set of 33 objects with the addition of I Zwl8[18,19]
NGC 5471[20] and CG 1116 + 51[21]. In addition changes have been made to
O/H and N/H for the SMC[20,22] and to NGC 2363[22,23] for O/H only.

There have been many attempts at finding a correlation between the
helium abundance and heavy elements. The first such attempt[24] was a
correlation of Y vs O/H. The simplest models of galactic chemical
evolution which assume that massive stars synthesize both Y and O leads to
the relation

$$Y_p = Y - Z \ dY/dZ \tag{3}$$
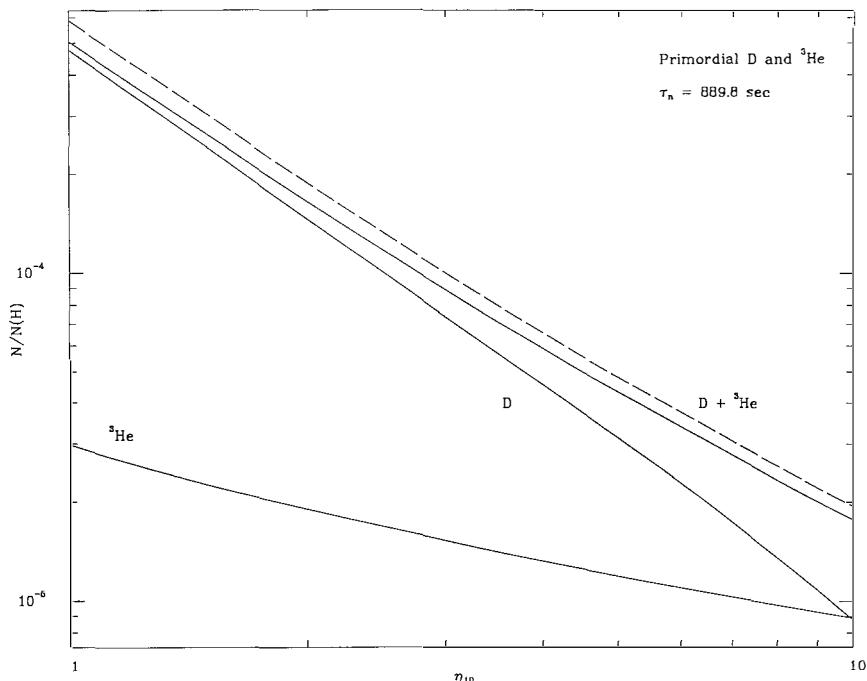
Figure 2

where Z is the metalicity which is well correlated to the oxygen abundance. Nevertheless, it stands to reason, that a more accurate determination of Y_p can be made by looking for low Z galaxies. Early investigations at trying to find a correlation lead to large variations in the fitted slope and the result depended somewhat strongly on which data points were included (most notably,the large Z values in Orion or errant low Y measurements of I Zwl8). A more recent survey of 12 metal poor galaxies found[25] effectively no correlation between Y and O/H. Kunth and Sargent[25] found that Y = .243 ± .010 + (35 ± 94)O/H but with a correlation coefficient r = 0.055. Instead, they argued that a weighted mean is more appropriate leading to the value Y_p = 0.245 ± 0.003. The Kunth and Sargent galaxies were included in Pagel's data set.

The data set for Y vs O/H is shown in Fig. 4.[13,14] A two parameter weighted fit to this data is,
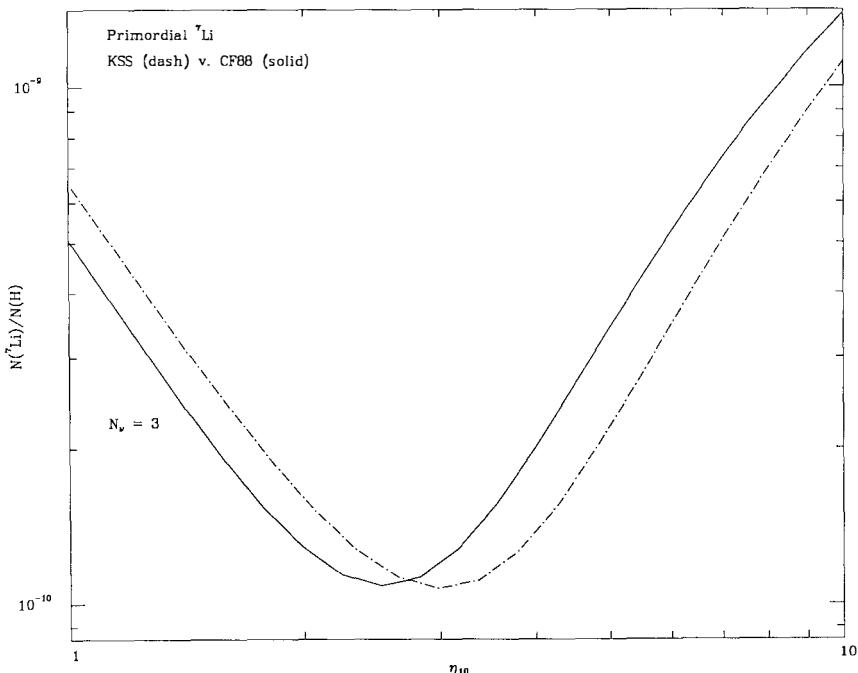
Figure 3

$$Y = 0.226 \pm .005 + (155 \pm 40)O/H. \tag{4}$$
$$r = .55$$

The errors in (4) are derived from the fit using quoted errors. The
correlation coefficient for this fit is also low, $r = .55$.

    I stress that great care must be taken in interpreting the
determined value of Y_p from (4). To begin with, Pagel's data sample has
included the collisional excitation correction but possibly in excess by a
factor of 2[17,26]. Furthermore, the uncertainties in the correction are
typically as large as the correction itself. It is also clear that
systematic errors have been neglected. Clearly there is reason to suspect
that (4) yields a value for Y_p which is too low and with underestimated
errors. In ref. 18, a careful study of IZw18 was made with special
attention to the errors. This is a particularly appealing galaxy since
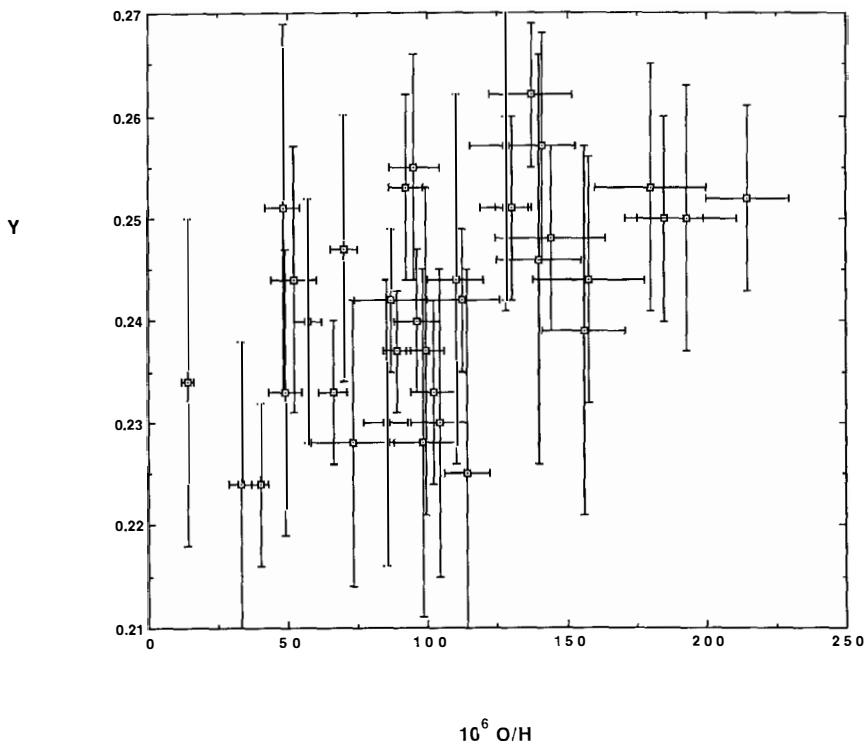
Figure 4

its oxygen abundance is the lowest available. However, because of the
larger errors quoted, this galaxy counts very little in a fit. If we take
the quoted error on IZw18[18] as a representative error, and take $\pm$ 0.016
in Y for all points, we get the fit

$$Y = .229 \pm .007 + (122 \pm 57)OH \qquad (5)$$

The fits of the data from the table are linear fits which extend
over 2 orders of magnitude in the oxygen abundance. There is really no
good reason to expect a linear relation over such an extended range.
Provided that we had sufficient data with low O/H, we could trust the
linear approximation. With only 36 data points to begin with, it might be

dangerous to chop off any high O/H galaxies. Nevertheless taking only the lowest 20 galaxies with O/H ≤ 105 x 10^{-6} and with equal errors in Y gives

$$Y = 0.232 \pm 0.010 + (68 \pm 139)O/H \qquad (6)$$
$$r = .19$$

and taking only the first 10 points with O/H < 90 x 10^{-6} gives

$$Y = 0.228 \pm 0.016 + (147 \pm 296)O/H \qquad (7)$$
$$r = .26$$

Notice the size of the "1-σ" uncertainty in the fit and the very poor correlation for low O/H galaxies. Though a poor correlation may indicate too short of a baseline, there is no reason to believe a linear fit over an extended baseline. What we need is more data!

It may not be surprising that we find such a poor correlation between Y and O/H. Indeed these elements are not made by the same stars. Oxygen is made by the more massive stars M ≳ 10 $M_\odot$, while most of the helium is made in intermediate mass stars. It has been argued that a better tracer of helium production in stars may be either nitrogen[15,17,26] or carbon[27]. The difficulty with carbon is that there are very few abundance measurements available. In Figure 5 the data for Y vs N/H is shown. Carbon gives a fit[13,14]

$$Y_p = 0.234 \pm 0.008 + (259 \pm 223)C/H \qquad (8)$$
$$r = .75$$

though with such little data, it is hard to draw strong conclusions from carbon other than that it is not in conflict with the other fits. For nitrogen, the data is fit to

$$Y_p = 0.230 \pm .004 - (2974 \pm 706)N/H \qquad (9)$$
$$r = .63$$

The nitrogen fit is somewhat more stable than the oxygen fit. Taking equal errors of 0.016 in helium for all points gives

$$Y_p = 0.231 \pm .005 + (2754 \pm 1135)N/H \qquad (10)$$

and for the first 20 objects with N/H ≤ 36 x 10^{-7}

$$Y_p = 0.226 \pm .009 + (5567 \pm 3751)N/H \qquad (11)$$
$$r = .51$$

and for the first 10 objects with $N/H \leq 25 \times 10^{-7}$

$$Y_p = 0.228 \pm .012 + (4010 \pm 7791)N/H \qquad (12)$$
$$r = .28$$

$$10^7 \text{ N/H}$$
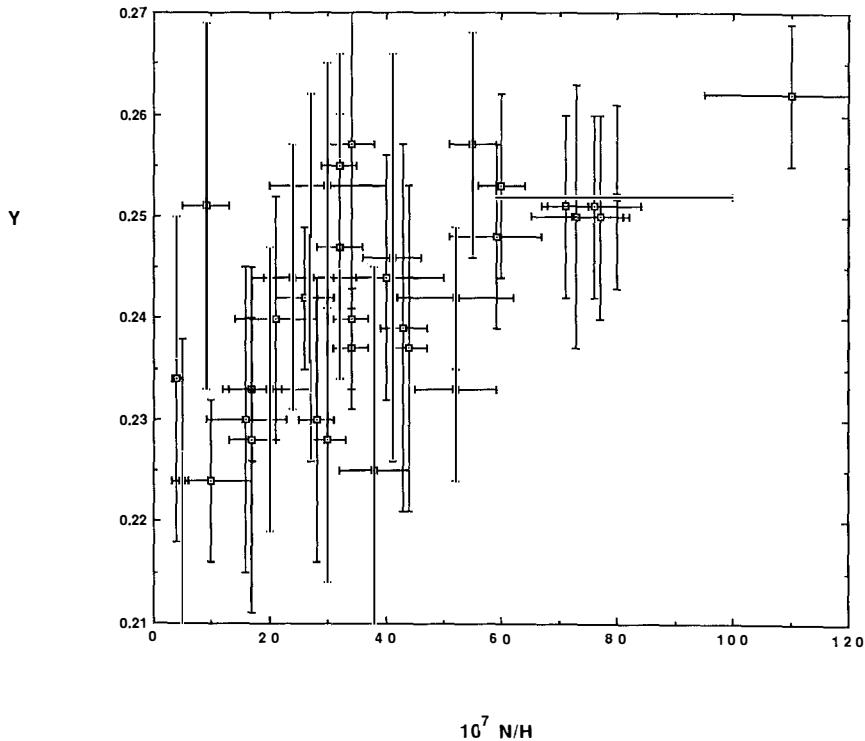
Figure 5

None of these fits show correlations which are satisfying, and it is difficult to draw too strong of a conclusion from this data . However, it is clear that the best value of primordial helium is below 0.25, and that data indicates a best value ≈ 0.23. The lack of stability of the previous fits has led us to conclude that a conservative estimate for Y_p is[13,14]

$$Y_p = 0.23 \pm 0.01 \tag{13}$$

[7]Li has been observed in halo dwarf stars, by several groups[28] giving population II abundances. In Figure 6, I show a compilation of this data, as a function for temperature for 35 stars with metallicity [Fe/H] $\leq$ -1.3. Units in the figure are such that [Fe/H] represents the logarithmic difference from solar values and [Li] = 12 + log([7]Li/H). What is markedly apparent in this figure is that there is a definite range T > 5500K and [Fe/H] $\leq$ -1.3 where the lithium abundance is very nearly constant. For these 35 stars one finds

$$[Li] = 2.08 \pm 0.12 \tag{14}$$

The error in (14) represents the scatter in the data. However, with the
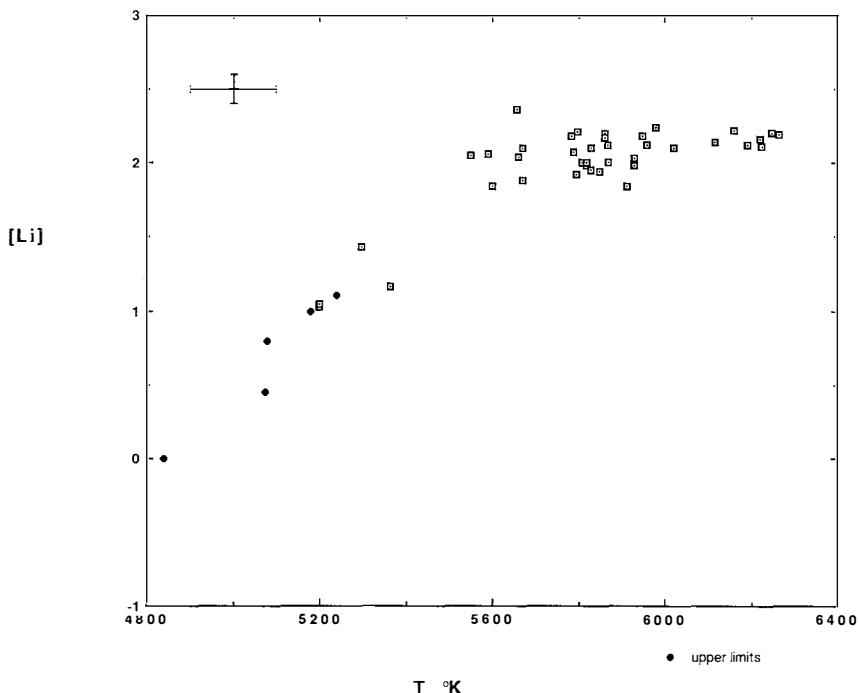
Figure 6

assumption that the pop II abundance is the primordial lithium abundance
then the error in the mean is more significant, in which case

$$[Li] = 2.08 \pm 0.035 \tag{15}$$

i.e., the mean value is quite well known. The existence of the plateau in
these figures is very good evidence that depletion[29] is not operative in
these halo stars and that these observations correspond to measurement of
primordial ^7Li. I should mention that in contrast to the case for ^4He,
the mean (15) is not sensitive to any particular object. In view of this
data, I would argue that there should in fact be no controversy with
regard to the ^7Li abundance. One should use the pop II abundance rather
than pop I.

From Figure 3, taking a 2-σ upper limit on $^7Li/H \leq 1.4 \times 10^{-10}$ from
Eq. (15), we find that η is tightly constrained

$$1.9 \times 10^{-10} \leq \eta \leq 3.4 \times 10^{-10} \tag{16}$$

Thus combining all of the limits on η from the light elements we
find[13,14]

$$2.8 \times 10^{-10} \leq \eta \leq 3.4 \times 10^{-10} \tag{17}$$

As one can see, this is a very small range for η ensuring consistency. A
central value of $\eta = 3.0 \times 10^{-10}$ implies the following light element
abundances:

$$D/H = 6.2 \times 10^{-5} \tag{18a}$$
$$^3He/H = 1.5 \times 10^{-5} \tag{18b}$$
$$Y_p = 0.237 \tag{18c}$$
$$^7Li/H = 1.3 \times 10^{-10} \tag{18d}$$

There is still an important consequence of the above limits. The
limit on η can be converted to a limit on the baryon density and Ω_B,

$$\Omega_B = 3.56 \times 10^7 \eta \; h_o^{-2} \; (T_o/2.7)^3 \tag{19}$$

and using the limits on η, Eq. (17), h_o and T_o from (2.7 - 2.8) K we find
a range for Ω_B

$$0.01 \leq \Omega_B \leq 0.085 \ . \tag{20}$$

Recall that for a closed Universe $\Omega > 1$, thus from Eq. (20) we can conclude that the Universe is not closed by baryons. Indeed, if $\Omega_B = 1$ and $h_o > 0.4$, $T_o < 2.8$ K we have $\eta > 4 \times 10^{-9}$ and the following light element abundances: $D/H \leq 10^{-7}$, $^3He/H \leq 5 \times 101^{-6}$, $^7Li/H \geq 5 \times 10^{-9}$ and $Y_p \gtrsim 0.269$ in clear contradiction with observations. This does not exclude the possibility that other forms of matter (e.g., massive neutrinos, etc.) exist in large quantities to provide for a large Ω. In fact, if $\Omega = 1$ as implied by inflation, the limit from nucleosynthesis would indicate that some form of dark matter must exist.

REFERENCES

1. J. Yang, M.S. Turner, G. Steigman, D.N. Schramm, and K.A. Olive, Ap. J. 281, 493 (1984).

2. A. Bosegard and G. Steigman, Ann. Rev. Astron. and Astrophys. 23, 319 (1985).

3. K.A. Olive, in the Proceedings of the International School on Astroparticle Physics, 2nd Course: The Early Universe and its Observable Consequences for Particle Physics, ed. D.U. Nanopoules, Erice, May 1989.

4. See e.g.,B. Adeva et al., Phys. Lett. B231, 509 (1989);D. Decamp, et al., Phys. Lett. B231, 519 (1989); M.Z. Akrawy et al., Phys. Lett. B231, 530 (1989); P. Aernio, Phys. Lett. B231, 539 (1989).

5. G. Steigman, K.A. Olive, and D.N. Schramm, Phys. Rev. Lett. 43, 239 (1979); K.A. Olive, D.N. Schramm, and G. Steigman, Nucl. Phys. B180, 497 (1981).

6. W. Mampe et. al., Phys. Rev. Lett. 63, 593 (1989).

7. G. Caughlan and W. Fowler, Atom. Dat. Nucl. Dat. Tal. 40, 29 (1988).

8. L. Kawano, D.N. Schramm, and G. Steigmann, Ap. J. 327, 750 (1988).

9. L.M. Krauss and P. Romanelli, see also L.M. Krauss, these proceedings.

10. J. Christensen et al., Phys. Rev. D5, 1628 (1972).

11. Particle Data Group, Phys. Lett. B (in press) 1990.

12. J. Byrne et al., Phys. Rev. Lett. (submitted) 1990.

13. K.A. Olive, D.N. Schramm, G. Steigman, and T. Walker, Phys. Lett. B 1990.

14. T. Walker, G. Steigman, D.N. Schramm, K.A. Olive, in preparation 1990.

15. B.E.J. Pagel, in "A Unified View of the Macro and Micro Cosmos," A. DeRujula, D. Nanopoulos, and P. Shaver, (eds.) (World Scientific - Singapore, 1989).

16. G.J. Ferland, Ap. J. 310, L67 (1986); R.E.S. Clegg, M.N.R.A.S. 229, 1P (1987); M. Peimbert and S. Torres-Peimbert, Rev. Mex. Astr. Astrof 14, 540 (1987) and 15, 117 (1987); R.E.S. Clegg, and J.P. Harrington, M.N.R.A.S., in press (1989).

17. B.E.J. Pagel and E.A. Simonson, Royal Greenwich Observatory preprint No. 104 (1989).

18. T.D. Kinman and K. Davidson, Ap. J. 243, 127 (1981); K. Davidson, T.D. Kinman, and S.D. Friedman, Astron. J. 97, 1591 (1989).

19. R.J. Dufour, D.R. Garrett, and G.A. Shields, Ap. J. 332, 752 (1988); E.D. Skillman, and R.C. Kenicutt, in preparation 1990.

20. S. Torres-Peimbert, M. Peimbert, and Fiero, Ap. J. (in press) 1989.

21. H. Dinerstein, G. Shields, Ap. J. 311, 45 (1986).

22. R.J. Dufour, F.H. Schiffer, and G.A. Shields, in Future of UV Astronomy Based on Six Years of Research, J.M. Mead, D. Chapman, and Y. Kondo, (eds.) NASA, CP-2349, p. 111.

23. M. Peimbert, M. Pena, and S. Torres-Peimbert, A.A. 158, 266 (1986).

24. M. Peimbert and S. Torres-Peimbert, Ap. J. 193, 327 (1974).

25. D. Kunth and W. Sargent, Ap. 273, 81 (1983).

26. B.E.J. Pagel, in Evolutionary Phenomena in Galaxies, J. Beckman and B.E.J. Pagel eds. (Cambridge University Press-Cambridge, 1989).

27. G. Steigman, J. Gallagher, and D.N. Schramm, Comm. Astrophys. (in press) 1989).

28. F. Spite and M. Spite, A.A. 115, 357 (1982); M. Spite, J.P. Maillard, and F. Spite, A.A. 141, 56 (1984); F. Spite and M. Spite, A.A. 163, 140 (1986); L.M. Hobbs and D.K. Duncan, Ap. J. 317, 796 (1987); R. Rebolo, P. Molaro, and J.W. Beckman, A.A. 192, 192 (1985); M. Spite, F. Spite, R.C. Peterson, and F.H. Chaffe Jr., A.A. 172, L9 (1987); R. Rebolo, J. Beckman, and P. Molaro. A.A. 172, L17 (1987); L.M. Hobbs and C. Pilachowski, Ap. J. 326, L23 (1988).

29. C.P. Deliyannis et al., Phys. Rev. Lett. 62, 1583 (1989).

BIG BANG NUCLEOSYNTHESIS: THE MORNING AFTER

LAWRENCE M. KRAUSS[1]

CENTER FOR THEORETICAL PHYSICS AND DEPT. OF ASTRONOMY

SLOANE LABORATORY, YALE UNIVERSITY

NEW HAVEN, CT 06511

Abstract

I review the status of primordial nucleosynthesis in the context of the standard homogenous big bang expansion, concentrating not only on the predictions, but the uncertainties.

INTRODUCTION

There are two fundamental sets of observations which form the empirical basis of big bang cosmology. The first is the observation of an isotropic thermal background of microwave radiation, which established both that the universe is expanding, and that matter and radiation were in thermal equilibrium at the surface of last scattering--when the radiation decoupled from matter around 100,000 years into the big bang. The second observation is that the abundance of Helium is everywhere around 25%. This fact suggests that our picture of the big bang expansion is accurate back at least until the universe was about 1 minute old. It is a direct prediction of standard big bang cosmology that about 1/4 of the universe by mass should have been converted around that time into Helium. Moreover, it appears impossible for so much Helium to have been created completely during the past 15 billion years of stellar manufacturing.

The past 15 years has seen the theory of big bang nucleosynthesis (BBN) go even further. It now appears that the inferred primordial abundance of all light elements up to and including Li can be accounted for consistently by BBN for a narrow range of cosmological parameters which happens to agree more or less wonderfully with the observational limits on these parameters. This agreement between BBN theory and observation caused a flush of excitement and confidence among cosmologists which has continued even up to the spectacular confirmation at the SLC and LEP, described during this meeting, that the number of light neutrino species is limited to be in more or less the same range that BBN predictions require it to be.

It is natural, of course, after the "rush" has worn off a little, to begin to examine the weak points instead of concentrating just on the positives. Indeed, the last few years has seen several alternative theories of primordial nucleosynthesis vying for our affections. These include the possibility that nucleosynthesis took place inhomogeneously, amidst a non-uniform distribution of baryons left over after the QCD phase transition in the early universe[1,2], and the possibility that a late stage of nucleosynthesis took place several tens of thousands of seconds after standard BBN, when hypothetical heavy unstable particles present in the universe may have decayed[3]. I will not describe in any detail these possibilities here. The first was briefly described by K. Olive and D. Schramm in their lectures during this meeting, and in any case, both possibilities appear to be severely constrained by observations. What I want to concentrate on here instead is to what extent we can assert that any of the predictions of BBN are what they are claimed to be. After all, if we are going to assert based on BBN predictions that the number of neutrinos is 3, or that the baryon density in the universe today is greater than the visible density of baryons, we had better be able to ascribe some confidence limits for the predictions themselves. It is only if a theory is robust enough to survive detailed investigations of its predictions that it is worth keeping around.

It is an opportune time to reconsider the predictions of BBN: (a) astronomical observations and theory are improving so that more detailed predictions of the model might be probed; (b) alternative models of primordial light element production described above have been explored which make predictions which marginally differ from those in the standard BBN model, and (c) experiments are providing ever better data to use in deriving BBN predictions. I therefore wish to

review here the standard BBN predictions, with an eye toward defining uncertainties and confidence limits. The direction of this work[4] is governed by several factors:

(1) The original BBN computer code of Wagoner[5] has formed the basis for most work in the field. Over the years a number of improvements, additions, and revisions have naturally taken place. Most recently, Kawano[6] undertook a major revision of the code which both updated and added reactions, and at the same time completely revised the user interface. This code is easy to use, and it is straightforward to vary the input parameters.

(2) As alluded to above, many of the reactions important for BBN have been remeasured since the original code was written, leading to revised estimates both of reaction rates and uncertainties. Most recently, new measurments of the following reactions have taken place: ^{3}He$(\alpha,\gamma)^7$Be, t$(\alpha,\gamma)^7$Li, ^{7}Li$(p,\alpha)^4$He, ^{7}Be$(n,p)^7$Li. In addition, the mean value of the neutron half life has been revised. This prompts one to examine the entire data set directly.

(3) Lithium has become an important discriminant in nucleosynthesis models. Alternative scenarios mentioned earlier predict either much more, or much less ^{7}Li, and most predict a significant abundance of ^{6}Li to be produced primordially. The BBN Li uncertainties, perhaps the largest of any of the light elements, therefore are essential to accurately estimate. A number of the new reaction rates quoted above can dramatically affect Li production.

With these factors in mind, one can perform a Monte Carlo analysis, simultaneously incorporating the experimental uncertainties associated with all of the most important imput reactions in the BBN code, to establish with some confidence the predictions and uncertainties of standard BBN. It is the results of this analysis[4] which I will review here.

First a caveat however. Perhaps the most important uncertainties relevant to any discussion of BBN are those based on observations which are used to infer the actual primordial abundance of light elements which one may compare the theoretical predictions. It is difficult to parametrize this observational uncertainty in any straightforward statistical manner, as should be clear from the talk by Olive at this meeting. It is an important prerequisite, however, when comparing theory and observation, to first get ones theoretical predictions straight, and I will review the standard BBN predictions and uncertainties here. Of course, I cannot resist making some provocative statements about the implications of these later, by comparing the predictions with various observational limits which have been bandied about, but I will draw no categorical conclusions in that regard here.

BIG BANG NUCLEOSYNTHESIS: THE HIGHLIGHTS

The following is a brief review of those features of primordial light element formation which will be relevant to further discussions here. The reader is referred to the literature[7] for a more complete discussion.

(1) As the early universe cools, the weak interactions which maintain a Boltzmann equilibrium distribution between protons and neutrons fall out of equilibrium. At this point, the neutron-proton ratio stops following its thermal equilibrium value n/p=exp ($-[m_n-m_p]/kT$), and freezes out at a value $\kappa \approx exp$ -$[m_n-m_p/T_{freeze\ out}]$. After this time, this ratio decreases very slowly due to free

neutron decay. The higher this ratio at the freeze-out time, i.e. the higher the freeze-out temperature, the more neutrons will be left to combine with protons to form nuclei heavier than hydrogen. The n/p freeze-out temperature depends upon two things---the expansion rate, and the strength of the weak interactions. The former depends on the square root of the number of relativistic degrees of freedom governing the expansion of the universe at the time of nucleosynthesis. Let us call the particles in the radiation gas, aside from $e^{\pm}$ and photons, N_v, to symbolize that this gas contains the known light neutrinos. However, it could also contain additional species, including other light, yet undiscovered particles. The greater the number of these, the faster the expansion rate, and the higher the temperature at which freeze-out occurs.

The strength of the weak interactions governing neutron-proton equilibrium is parametrized by the neutron half life, $\tau_{1/2}$. If $\tau_{1/2}$ is increased, the weak interaction rates are decreased, and n/p freeze-out occurs at a higher value. **The empirical success of BBN is based upon the fact that, given the approximate strength of the weak interactions and the expected number of particles in the radiation gas at at temperature of about 1 MeV, κ can be estimated to be about 1/7.** Since virtually all free neutrons which survive will later be captured to form ^{4}He, the predicted fraction by weight of Helium compared to all baryons in the universe is $\approx 2\kappa/[\kappa+1] = 1/4$, in good qualitative agreement with all observations. Whenever the freezeout temperature is increased, so that the neutron to proton ratio at freeze out is increased, the prime beneficiary is ^{4}He. Nevertheless, the abundance of other trace elements, deuterium and lithium for example, will also be affected.

(2) The efficiency with which neutrons will be captured to form ^{4}He depends also upon one other parameter, the number density of protons and neutrons relative to photons in the universe at the time of nucleosynthesis. This can be expressed in terms of the number density of baryons today, which can in turn be expressed in terms of the density of photons today in the observed cosmic microwave background. Thus, one defines the quantity η, the ratio of baryons to photons in the universe at the time of nucleosynthesis. If one assumes no significant non-adiabatic reheating processes since nucleosynthesis then one obtains the following relation between η and Ω_B, the present baryon mass density as a fraction of the closure density, and the present temperature of the microwave background T_0:

$$\Omega_B = .0036 h_o^{-2} (T_0 / 2.74 \text{ K})^3 \eta_{10} \qquad (1)$$

where $\eta_{10} = 10^{10} \eta$, and the Hubble constant is $H_o = 100 h_o \text{ kms}^{-1}\text{Mpc}^{-1}$.

One might imagine that increasing the baryon density at the time of nucleosynthesis would keep the weak interactions in equilibrium longer, and result in a lower freeze-out temperature, and hence a smaller remnant ^{4}He abundance. However, another effect works more efficiently in the opposite direction. Because the binding energy of deuterium and heavier nuclei is O(MeV) no appreciable deuterium forms until the universe cools to a temperature of about 0.1 MeV. Only after this "bottleneck" is passed can the formation of heavier elements proceed. As the baryonic abundance is increased, the formation of deuterium begins slightly earlier, when the density of

matter and radiation is higher, and the efficiency of nuclear burning to helium is greater. This results in more helium, and less remnant deuterium (and ^{3}He, which is also produced during the process of burning to ^{4}He). Thus, the helium fraction is a monotonically increasing function of η, while the D and ^{3}He fractions are monotonically decreasing functions of η.

(3) The mass gaps at atomic mass 5 and 8 (no stable nuclei) hinder the production of nuclei more massive than helium by reactions of helium with itself, or with hydrogen. ^{7}Li is the heaviest of the stable light elements formed with an abundance of greater than one part in 10^{10} compared to hydrogen in the big bang. Because the ^{7}Li remnant abundance is so small, and because it is produced after helium, it is more sensitive than any of the other light elements to the detailed nuclear reaction dynamics in BBN, and hence is more sensitive to any present uncertainties in our knowledge of the rates of these reactions. There is one feature of its predicted abundance which is notable however. Alone among the light elements, the predicted ^{7}Li abundance has a minimum as a function of η, and moreover this minimum occurs in a physically interesting range. The presence of this minimum is easily understood on physical grounds. Because of the increasing efficiency of Li destruction, via the reaction ^{7}Li(p, α)^{4}He, with increasing η, the remnant abundance of directly produced ^{7}Li falls with η. However, ^{7}Li is also produced after BBN via the decay of the unstable isotope ^{7}Be. This isotope is produced with increasing efficiency as η increases, until its remnant abundance (before it decays) exceeds that of the directly produced ^{7}Li abundance. Thus, for low values of η the ^{7}Li abundance falls with η, but at a certain point the ^{7}Li which is produced indirectly by the decay of ^{7}Be begins to dominate, and the remnant ^{7}Li abundance then begins to grow with η. This implies that increases in reactions which create (destroy) Li increase (decrease) its abundance for low values of η, while increases in those which create (destroy) Be increase (decrease) the Li abundance at high η.

REACTION RATES AND UNCERTAINTIES:

Since the initial calculations of BBN, physicists have worked to determine the values of reaction rates which are of importance in both primordial and stellar thermonuclear processes. In the past twenty years, many of the reaction rates have been remeasured, and at the same time, theoretical estimates have improved. It has been a common practice when making BBN estimates to re-evaluate primordial abundances using only the most recent experimental and theoretical values. This can lead to misleading results. First, it is not clear that more recent experiments always lead to improved values. Second, it is not clear from this kind of analysis how systematic and statistical errors might affect the results. Even an approach based on incorporating uncertainties in individual reactions does not allow one to place confidence limits on BBN estimates. Because different reactions play a different role for different values of η, errors need not necessarily combine in any straightforward fashion. Finally, unless error estimates are tied to a direct analysis of the data one cannot accurately gauge the margin for improvement.

In all, we incorporated experimental uncertainties for ten of the major BBN reactions, including neutron decay, and directly analyzed the experimental data for those reactions which have

recently been remeasured. We concentrated initially on the reactions related to lithium production for several reasons. Besides the renewed importance of lithium as a discriminant for various BBN scenarios, and the recent work on deducing the actual primordial abundances of ^{7}Li and ^{6}Li [8], the largest uncertainties are associated with these reactions, and as a result, recent re-measurements of several key reactions have been undertaken. In certain cases it was clear that systematic errors exceed statistical ones, and we dealt with these on a case by case basis. In addition, we also utilized quoted uncertainties in reactions which have not been re-measured recently, or for which we could not directly access the data. In all the reactions considered except three, the central values used for reaction rates are those quoted in the previous literature (up to small ($<10\%$) revisions in certain cases which we have checked would leave our quantitative results essentially unchanged).

In general, experiments at best involve a measurement of $\sigma(E)$, the total cross section at center of mass energy E. From this, the low energy astrophysical S factor is calculated :

$$S(E) = \sigma(E) \, E \exp(2\pi Z_1 Z_2 \, e^2/v) \tag{2}$$

where $Z_1 Z_2$ is the product of the atomic numbers of the interaction particles, and v is their relative velocity. This value is essentially the cross section with the Coulomb barrier removed, and is much better behaved at lower energies than $\sigma(E)$. $S(E)$ is frequently expressed as $S(0)$ multiplying a Taylor expansion in E:

$$S(E) = S(0) \{ 1 + (S'(0)/S(0)) \, E + 1/2(S''(0)/S(0))E^2 + \ldots \} \tag{3}$$

$S(0)$ is also occassionally referred to as the astrophysical S factor. The more relevant quantity for BBN is the reaction rate as a function of temperature T. This can be found as a power series in T from the expansion for $S(E)$ by convolving the relevant quantities with the thermal distribution functions for the particles involved. Such rates are frequently expressed as $S(0)$ multiplying a function of T. The uncertainty in $S(0)$ is used to parametrize the uncertainty in the entire rate.

In an ideal case one would consider reaction rates with uncertainties that vary as a function of energy, or temperature. As a first step however, one can assume the reaction rate as a function of energy as known (this is what is used to derive $S(0)$ in the first place), and then parametrize the effect of the measured uncertainties at different values of energy in a single fractional uncertainty in $S(0)$. This of course can be a dangerous procedure, since the uncertainties which contribute most to $S(0)$ may be those which are irrelevant to the reaction rate at the energies of interest in BBN. In particular, low-energy resonances which might alter $S(0)$ need not affect the behavior of $S(E)$ in the energy range of interest for BBN. Indeed, this kind of mis-identification of uncertainties has appeared in some of the previous literature. When the uncertainties vary strongly with energy, we have used *the largest uncertainty in the relevant energy range for BBN* to characterize the uncertainty of the rate, parametrized by a quoted uncertainty in $S(0)$. In fact there are several cases where if this is not done, the uncertainty in $S(0)$ would lead to a mis-estimate of the actual uncertainty relevant for BBN considerations. In any case, quoting the uncertainty this way, in terms of $S(0)$, gives a conservative estimate that is both physically reasonable, and easy to implement in the computer code.

Individual Reaction Analysis

(a) $^3He(\alpha,\gamma)^7Be$: This reaction has been one of the most carefully studied reactions because of its importance in the solar production of ^8B neutrinos. There is a good theoretical understanding of the reaction using resonating group calculations[9], as well as a wide variety of measurements for it[10]. The results of these are shown in figure 1 below, displaying S(0) for the 8 measurements.

Fig. 1
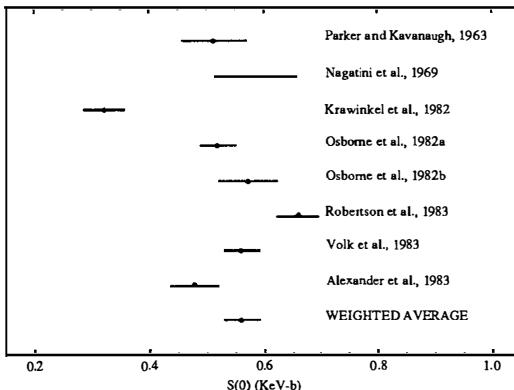
There is good agreement among seven of them, while the Krawinkel et al. 1982 data is anomalously low. We used the seven concordant results to find an average value, and increased the uncertainty in this average value slightly, compared to the statistically derived 1σ value, to obtain a 1σ overlap between all experimentally measured points and the average. We note that a factor of two uncertainty has been quoted for this reaction in the past, due to the inclusion of the Krawinkel data in the calculation of the rate. However, as the averaged data clearly show the 1σ uncertainty is much more like 6%. Note also that (a) because there is good agreement between the shape of the theoretical and experimental curves, and (b) because the expermental errors do not vary strongly with energy, the parametrization of the energy dependent error in terms of a single quoted normalization error on S(0)---obtained by extrapolating a fit of the theoretical curve to the data--- seems a good approximation in this case. Fig. 2 displays the effect on BBN ^7Li production (expressed as a fraction of H by number) of varying just this reaction within its 1σ and 2σ limits. As can be seen, increasing the reaction rate increases the Li abundance at high η, where essentially all of the remnant Li abundance is due to decay of Be.

Fig. 2: $^3\text{He}(\alpha,\gamma)^7\text{Be}$
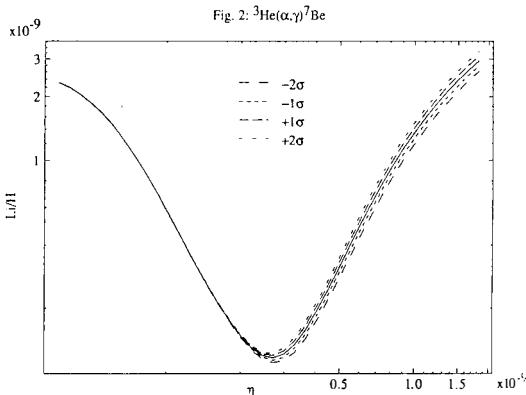
(b) $t(\alpha,\gamma)^7\text{Li}$: This is the mirror reaction to the first reaction. As a result it is also reasonably well understood theoretically[11], but is less well measured because of experimental difficulties in dealing with tritium. Because of the theoretical advances, and because of new measurements, this reaction rate has altered more since BBN estimates began than perhaps any other. Original estimates used a straight line extrapolation from the measured lowest S(E) value to derive an S(0) value of about 0.06 kev-barn[12]. However, using the mirror reaction data and theory, a theoretical curve going through the same data would raise this S(0) value to about 0.1 kev-barn. Beyond this, the experimental uncertainties are such that a recent re-measurement by one group appeared to alter S(0) by almost 50% even above this higher value. There now exist three different measurements of this reaction, not all of which are in agreement. Following ref. 10 we took the theoretical curve for S(E) derived from the mirror reaction and normalized it to each of the three data sets, with S(0) as the free parameter, using a least squares fitting technique. We then found the weighted average of S(0), and took as our uncertainty a value that incorporated all the data, which turns out to lead to a 1σ uncertainty of about 17%. This is shown in Fig. 3.

Fig. 3
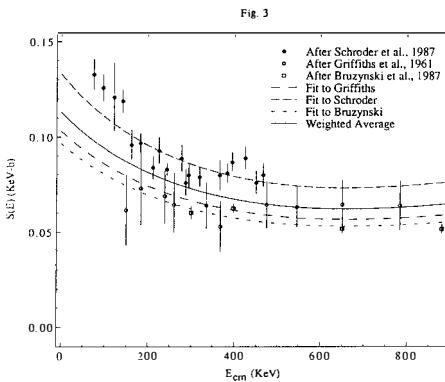
This uncertainty is larger than the statistical uncertainty one would derive, but in the presence of measurements which are in clear disagreement, until one can argue convincingly that one of the data sets should be discarded, this seems all that one can do. (It is likely that further re-measurements will confirm the lower values, leading one to discard the anomalously high data set (W. Fowler, private communication).) Also note again that because the shape of the theoretical curve is fixed, and the errors which govern the curve fitting are not grossly energy dependent, the uncertainty in $S(0)$ accurately reflects the uncertainty in the rate in the energy range of interest for BBN. Fig. 4 displays the effect on BBN ^{7}Li production of varying this reaction within its 1σ and 2σ limits. Here, reducing the rate has the effect of reducing the remnant Li abundance for low η.

Fig. 4: ^{3}H$(\alpha,\gamma)^7$Li
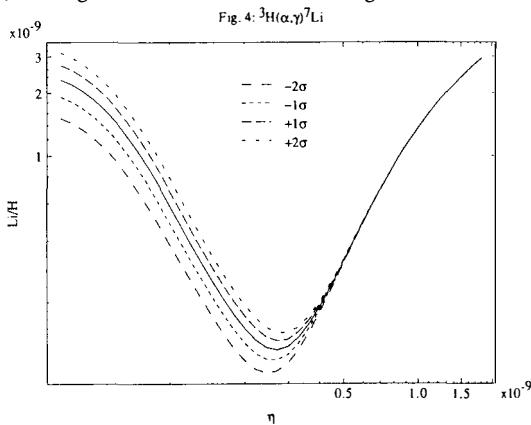
(c) ^{7}Li$(p,\alpha)^4$He: This is the major lithium destruction reaction in stellar interiors and in BBN. Rolfs and Kavanaugh (RK)[13] have recently remeasured the cross-section, and have found a substantially lower value of $S(0)$ than previously estimated. However, this is in large part due to new, more accurate measurements for energies less than about 100 keV, for which $S(E)$ appears to drop faster than previously estimated. For energies in excess of this value, $S(E)$ is unaltered. We compared the RK value of $S(E)$ with the previously determined experimental values, and found the new measurement of $S(E)$ to be in good agreement with previous experiments. Thus, the reduced value of $S(0)$ does not reflect a change in the behavior of $S(E)$ in the range of interest for BBN. Moreover, the 16% error quoted on $S(0)$ by RK is not appropriate for use in BBN estimates. This large error is due to extrapolating curves through the poorly measured points at low energy. At higher energies, relevant for BBN, the data is much better determined, with an uncertainty of 8%, coming from an overall uncertainty in the absolute cross-section normalization. It is this latter uncertainty that we used in our estimates. Fig. 5 displays the effect on ^{7}Li of varying this reaction by 1 and 2σ. As expected, increasing this Li destruction reaction rate reduces the remnant Li abundance for low values of η, where the remnant abundance is primarily due to directly produced Li, and not due to the decay of primordially produced Be.

112
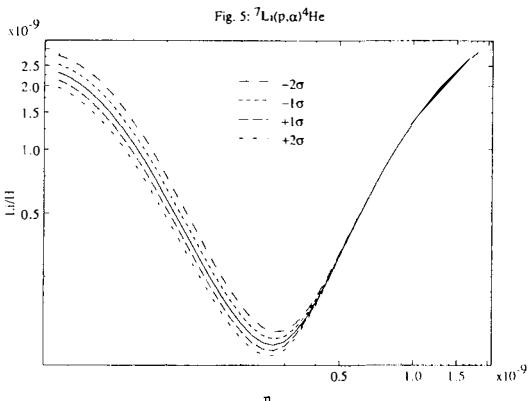
Fig. 5: $^7\text{Li}(p,\alpha)^4\text{He}$

(d) $^7\text{Be}(n,p)^7\text{Li}$: This reaction is important for determining the amount of Be which is converted into Li before the end of BBN. Since Be is less susceptible to destruction during BBN, an increase in this reaction rate results in a smaller remnant abundance of Li, since less remnant Be remains to decay into Li after BBN has completed. For this reaction there appears to have been a great confusion about uncertainties, with uncertainties quoted between 20% and a factor of 2. Our initial work on Li abundances[14] utilized the reaction rate values and uncertainties quoted by Bahcall and Fowler (BF)[15]. However, recently this reaction has been remeasured, leading to an alteration in the value of S(0) compared to that utilized in the earlier BF analysis. We re-analyzed the data and included the updated values, leading to some slight alterations compared to the numbers quoted in ref 14. The major change in the measurements of the $^7\text{Be}(n,p)^7\text{Li}$ reaction have to do with the cross sections for thermal neutrons. Traditionally the reverse (p,n) reaction was measured and reciprocity relations were used to derive the (n,p) reaction rate. However, recently a Los-Alamos group[16] has directly measured the neutron induced reaction rate, and found the thermal cross section to be almost 25% smaller than the earlier quoted value, with an accuracy of 10 times that of the original measurements. The thermal neutron cross section is not directly relevant for BBN, however. One has to determine how these new measurements affect S(E) in the region of several tens to several hundred keV. In this region the new data is in fact in good agreement with the older data, with uncertainties which are on same order of magnitude as the uncertainties in the older measurements, i.e. there has been no improvement in the accuracy with which this cross section is known in the relevant energy range. One might expect this to imply that no change is required in the overall value of S(E), or in its uncertainty in a BBN analysis. This is not the case however. BF calculated a theoretical curve for this reaction which, while it incorporated certain of the higher energy data, was normalized at low energies by the old S(0) value. Thus, when one compares their curve with the data one finds that it overestimates S(E) at all energies. We have fit a new curve for $<\sigma v>$, the rate as a function of temperature, which is inferred from an S(E) curve which goes through the center of all of the data at high energy (and the

new data at low energy). This leads to a lower value of <σv> (by about 20%) than previously utilized. Our expression for <σv>, and hence the shape of the curve we use, is determined by fitting the coefficients in the same polynomial expansion with temperature previously utilized in the Kawano code for this reaction, to the <σv> curve we determined by fitting the data.

Fig. 6
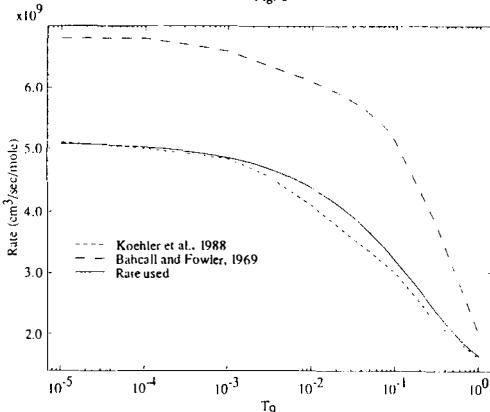
The expression we derive is:

$$N_A <\sigma v> = 5.1 \ (1 - 1.515 \ T_9^{1/2} + 1.173 \ T_9 - 0.336 \ T_9^{3/2}) \qquad (4)$$

where N_A is Avogadro's number and $T_9 = T/10^9$ K. The uncertainty we estimate in <σv> is given by the variance of the data, in the energy range of interest for BBN, determined by eye from fitting curves to the new data. It is approximately 10% and is uniform across this energy range. Thus, it is appropriate to parametrize this uncertainty by an overall multiplicative constant in the expression given above. Fig. 7 displays the effect of varying this reaction within its 1σ and 2σ ranges on Li production in BBN. As expected, the major effect is for high η, where essentially all the remnant Li comes from Be decay. Decreasing this reaction increases the remnant Li abundance at high η.

Fig. 7: $^7Be(n,p)^7Li$
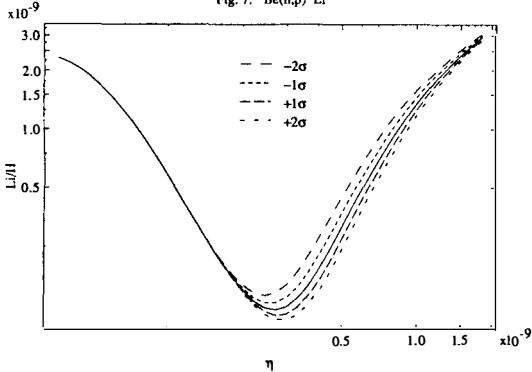
(e) Neutron Lifetime, τ: This is the parameter which affects the overall remnant ^4He

abundance today arising from BBN, as we described earlier. There has been an important change in this parameter since most earlier analysis of BBN uncertainties and predictions[4,14]. Namely, the measured half life, $\tau_{1/2}$, has decreased dramatically compared to the value quoted 5 years ago. From a particle physics perspective, this decrease in the neutron half life, which implies a higher g_A/g_V ratio for weak interaction parameters, now brings the measured value for this ratio from neutron decay into agreement with the value obtained from other weak interaction measurements such as hyperon beta decay. Previously this important ratio, determined from other experiments to be about 1.259, exceeded the value obtained from neutron half life measurements. In any case, aside from the importance of this measurement for particle physics, decreasing the neutron lifetime, and increasing g_A/g_V also implies that weak interactions stay in equilibrium longer during BBN, which has the effect of reducing the remnant ^{4}He abundance. Since this is one of the most significant factors leading to quoted bounds on the number of neutrinos (or relativistic degrees of freedom), it is important to examine the new neutron half life data in any BBN analysis. Fig. 8 presents the results from all experiments performed at the time this analysis was begun. The value we obtained was $\tau = 900 \pm 8.6$ sec ($\tau_{1/2} = 624 \pm 6$ sec)[4].

Fig. 8

Sosnovskii et al., 1959
Christensen et al., 1972
Krohn and Ringo, 1975
Bondarenko et al., 1978
Stratowa et al., 1978
Erozolimskii et al., 1979
Byrne et al., 1979
Bopp et al., 1986
Last et al., 1988
WEIGHTED AVERAGE
Particle Data Tables, 1988

550   600   650   700
Neutron Half-life (sec)
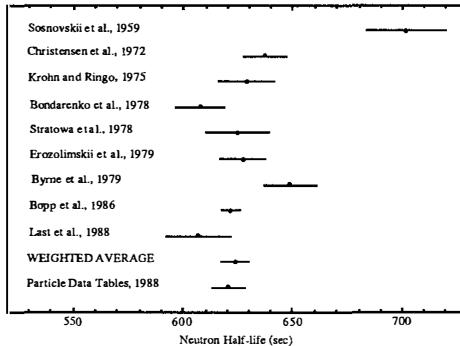
At the same time, recalculations of τ utilizing both the neutron decay data and other particle physics data were performed to update the value of τ for the tables in the Review of Particle Properties 1988[17]. The mean value and uncertainty quoted there is $\tau = 896 \pm 10$ sec($\tau_{1/2} = 621 \pm 7$ sec). This value, which matched very well with our previous analysis, is what we chose to use. Most recently, however, an experimental result appeared yielding a new, lower and apparently more accurate measurement of $\tau = 887.6 \pm 3$ sec ($\tau_{1/2} = 615 \pm 2$ sec)[18]. If it were amalgamated in a straightforward statistical fashion, it would dominate in the determination of the neutron mean lifetime and uncertainty. However, since this measurement does not overlap with some existing data, it is not clear whether or not the error should be increased over its strict statistical values when quoting a conservative value for the uncertainty. In any case, even incorporating this new result directly would not greatly change our results. As will be described in a later section, what is

most important for placing an upper limit to the number of relativistic degrees of freedom existing during nucleosynthesis is the lowest possible value of τ. The 2σ lower value for $\tau_{1/2}$ we utilize is 607 seconds, while that appropriate to this recent measurement is 611 seconds. The difference is less than 30% of the net 2σ variation we consider.

Fig. 9 displays the effect of varying the neutron lifetime within a 1σ and 2σ uncertainty on the remant primordial mass fraction of ^{4}He. For comparison purposes, we show in Fig. 10 the effect of the same variation, using only the Mampe et al value for the mean lifetime and uncertainty.

Fig.9: $\tau_{1/2}$
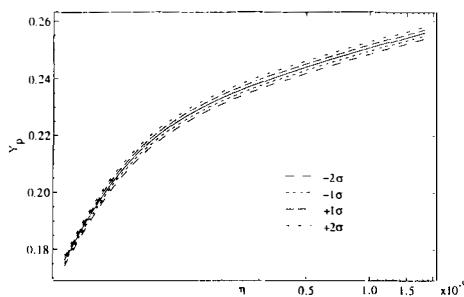
Fig. 10: $\tau_{1/2}$ from Mampe et al 1989
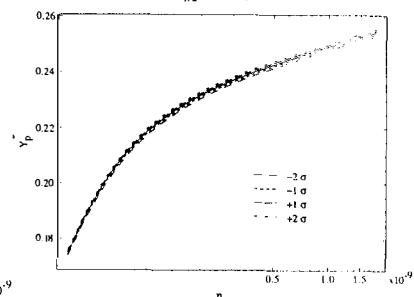
(f) Other Reactions: For the remaining reactions we utilized the mean values and uncertainties quoted in earlier work. For the reaction d(p,γ)^{3}He we used the analysis of Bahcall et al[19] who quote an uncertainty in S(0) of 16%. For the remaining 4 reactions, p(n,γ)d, d(d,n)^{3}He, d(d,p)t, and t(d,n)^{4}He, the 10% uncertainties quoted in ref 20 were used. The remnant Li abundance behaves as one might naively expect for all reactions (i.e. increasing a reaction rate for a reaction with an end product which contributes directly to Li production increases the remnant abundance for low η, while increasing a reaction rate whose product participates in a Be producing reaction increases Li abundance at high η) with the exception of the p(n,γ)d reaction. Here a much more complicated behavior is observed, reminding us that several competing factors are relevant when considering the effects of initial BBN reactions. For example, increasing the p(n,γ)d rate decreases the remnant deuterium abundance as a function of η. This is because while this reaction produces deuterium, increasing the rate causes deuterium to be produced earlier, when it is then burned more efficiently. What remains surprising, however is the sensitivity of Li to this initial BBN reaction. A 10% variation in this reaction changes the remnant Li abundance by as much as 30%.

MONTE CARLO ALGORITHM:

A Monte Carlo analysis was performed to take into account simultaneously all uncertainties in reaction rates and in the neutron half-life, each of which enters in a complicated way in determining the overall abundance of light elements produced in BBN. This required generating individual reaction rates with a gaussian distribution about the mean value, and with width determined from the uncertainty analysis described above. The 1σ uncertainties in the individual reaction rates were taken to represent overall uncertainties in each rate. That is, a 10% 1σ uncertainty was taken to

mean there is a 66% probability for the entire rate to be between 10% higher than and 10% lower than the mean, at each temperature. This allowed us to incorporate uncertainties directly in the code in the form of randomly generated prefactors. For each rate, the multiplicative prefactors were generated from a gaussian distribution, centered around 1, with width given by the 1σ uncertainty in the rate. These prefactors then multiplied the entire rate, whose overall form was identical with that in ref. 6, except in the cases explicitly described earlier. The gaussian distributions were truncated so that all prefactors were positive definite, although this wasn't needed in a statistically significant way - since the largest uncertainty used was 20%, generating a number less than 0 with this procedure is a 5σ event.

The procedure in terms of programming was straightforward. The initial value of η and the number of neutrinos were fixed. The program then generated a random prefactor for each rate, and for each such set of rates the remnant light element abundance was determined. We repeated this procedure 1000 times, for each of 20 values of η, and for 3, 4 and 5 neutrino families. Our choice of values of η was made to best follow the structure of the Li vs η curve, which is the only one which is not monotonic. The 1σ curves were defined to bracket the central 66% of the final predicted abundances, for each value of η and N_ν, and 2σ curves bracketed the central 95%. This Monte Carlo procedure required approximately 1000 hours of CPU time on a VAX 11-780.

RESULTS AND DISCUSSION:

The results of the analysis described here can be presented succinctly. Figs. 11-12 display the combined 1 and 2σ limit curves, described above, for various key "observable" light element abundances: ^{4}He/H by weight, and ^{7}Li/H, D/H, (D+^{3}He)/H, and ^{7}Li/D by number. The D and ^{3}He limits are presented in this form because the combination is claimed to be more easily constrained by observation than either isotopic abundance separately[20].

Fig. 11   Fig. 12
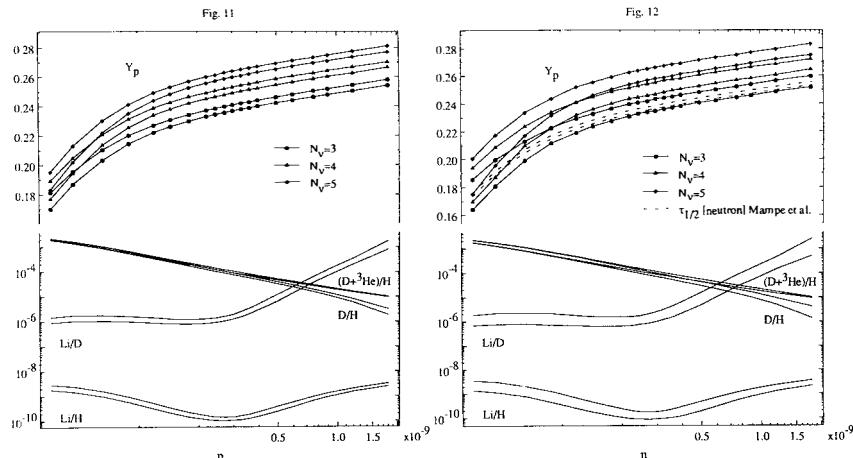
Even a brief inspection of Figs. 11-12 provokes two remarks. First, **the "predictions" of standard BBN must be represented by "bands" rather than lines on an abundance versus η curve**. This is because the presently existing experimental uncertainties are sufficiently large so that they cannot be ignored when attempting to derive bounds from BBN on various fundamental cosmological and particle physics parameters. Second, *in spite* of these experimental uncertainties, there remains a restricted range (albeit now somewhat less restricted) over which a general agreement of predictions and observations is possible. As a result, bounds can clearly be derived on quantities such as η, and the number of relativistic degrees of freedom present at temperatures of about 1 MeV. The advantage of using limits derived from the experimental uncertainties in a statistically consistent manner, however, is that one then can associate for the first time realistic confidence limits with one's predictions.

Of course, as I have stressed, when comparing BBN predictions and observations, probably the greatest and certainly the least quantifiable uncertainties are associated with the observations. Here the possible sources of error include not only those directly associated with the observations themselves, but more importantly, those associated with the use of present observed isotopic abundances to infer primordial values. Even in the case of the best measured element, ^{4}He, there is not yet a statistically clear relationship between isotopic abundance and metallicity which allows a unique primordial value for He to be extracted with high confidence (see K. Olive's talk). I do not have the space here to attempt to review the observational situation in detail, and therefore cannot quote categorical limits here. Nevertheless, it seems useful to point out what the implications of the BBN predictions are for cosmology and particle physics, as a function of various ranges one might hope to place on light element abundances from observation, and to examine possibilities for the future.

Let us begin with a lower limit which might be placed on η, the baryon to photon ratio, and hence on the present density of baryonic matter in the universe. Extrapolating back to determine the remnant abundance of D from present observations is subject to large uncertainties, because of the ease with which deuterium is destroyed in stars. However because D is burned to ^{3}He in stars, with the latter being more difficult to destroy, the remnant abundance of $(D+^3He)/H$ is claimed to give a more reliable observational upper limit which one can compare to predictions[20]. Using a "safe" observational upper limit of 10^{-4} for this ratio[20], shown in Fig. 13a, this crosses our 2σ curve at $\eta_{10}=2.6$. This lower bound on η is significantly higher than that which may be obtained from the other light element abundances. If one chooses not to use it, then requiring that the primordial abundance of ^{4}He $\geq 22\%$, which seems a conservative assumption (see below), implies $\eta_{10} \geq 0.8$ (for $N_\nu \leq 3$), while the limit obtainable using ^{7}Li (see below) is closer to $\eta_{10} \geq 1.2$. (It should be noted that even this lower bound, when combined with lower limits on the age of the universe, is marginally suggestive of the possible existence of at least some baryonic dark matter.)

Fig. 13a
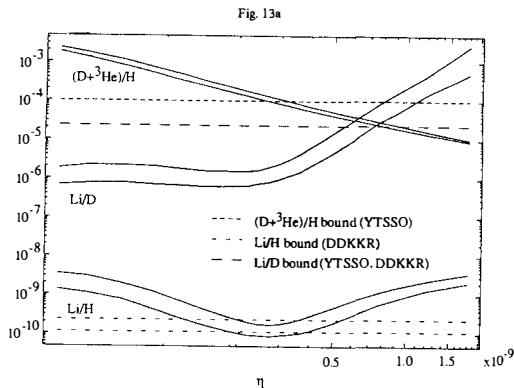
The lower bound on η obtained from $(D+{}^3He)/H$, has been combined with limits on ${}^4He$ both to check for consistency of standard BBN, and to derive well known limits on the number of light relativistic degrees of freedom during BBN, including light neutrinos. From Figs. 11-12 it can be seen that an upper limit on primordial ${}^4He$ can be used to set upper limits on η, and more importantly on the number of light particles in abundance at the time of nucleosynthesis. As a result, a great deal of effort has gone into trying to determine whether more than three light neutrino species might be consistent with BBN, or whether a fourth light family might be ruled out by cosmology before it could be directly probed at accelerators.

This issue may now seem moot. As described at this meeting, Mark II at SLC and four experiments at LEP have now clearly established that four light neutrinos are ruled out by direct measurements performed at the Z^0 peak[21]. However it should be stated that the BBN bound and the Z^0 limit are in some sense complementary. In particular while the Z decay measurement is sensitive to heavy neutrinos which might not play a role during BBN[22], it should be pointed out that BBN is sensitive to light degrees of freedom which can decouple from the Z. Hence, it is altogether non-trivial to continue to examine the question of whether more than effectively three neutrino species might be compatible with BBN predictions.

As described earlier, the measured neutron half life has decreased considerably in the past several years. This reduces the predicted remnant ${}^4He$ abundance from BBN. In addition, the inclusion of the uncertainties in the predicted $(D+{}^3He)/H$ abundance lowers the lower bound which might be quoted on η. This also has the effect of relaxing the BBN bounds on the number of "neutrinos". What is the result? Because of the importance of this limit, and its dependence on the ${}^4He$ abundance estimates which we cannot adequately critique, we prefer to present several scenarios, depending upon one's choice of upper limit for ${}^4He$. In Fig. 19b we show the ${}^4He$ 2σ limit curves, and five lines, corresponding to 23.5%, 24%, 24.5%, 25%, and 25.5% respectively.

Fig. 13b
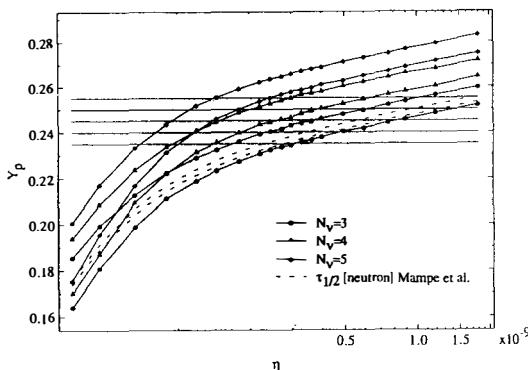
If one were to choose the latter, quoted in ref. 20 as a very conservative "3σ" limit, then, using the reduced lower limit on η quoted above, five "neutrino" degrees of freedom would just be marginally ruled out at the 2σ level. However, as recently stressed, (i.e. see K. Olive, this meeting) the measurements of ^{4}He have improved in the intervening years, and it is likely that a lower value should be used. If one reduces the upper limit on ^{4}He to 25% and uses the bound on η quoted above, then five "neutrinos" is clearly ruled out, but four "neutrinos" is easily consistent. If one reduces the value to 24.5%, which was suggested by Olive and Schramm here as a reasonable present upper limit, then four "neutrinos" is again just marginally acceptable, but only at *both* the (D+^{3}He)/H and ^{4}He 2σ extreme limits. Finally, if one puts an upper limit of 24%, then not only is a fourth neutrino-like degree of freedom ruled out, but *there is only a narrow range of η where even three neutrinos is acceptable.* In particular, if one were to use the lower bound on $η_{10} \geq 2.6$ from (D+^{3}He) quoted earlier, and the new neutron half-life experimental result then *any value of* ^4He <23.5% *would be inconsistent with BBN with 3 neutrino species.* We shall have more to say about this later.

Finally, regarding cosmological limits on the possible consistency of BBN predictions and observations, what about an upper bound on η? Up till now, this has been provided by estimates of either ^{7}Li/H or ^{7}Li/D compared to observational limits. While both the BBN uncertainties and the observational uncertainties are greatest for Li among all the light elements, there has been a great deal of progress in both areas in recent years. Nevertheless, upper limits on ^{7}Li are controversial. Recently it was argued [14] that the halo star abundance could be well fit by a restricted range of primordial abundance, $1.09 \times 10^{-10} < ^7Li/H < 2.3 \times 10^{-10}$, thus allowing one to place a new upper limit on η by combining this limit with our uncertainty analysis. In the meantime there have been several new developments. First, the BBN estimates have been updated by using the new ^{7}Be(n,p)^{7}Li data [4]. This has the effect of reducing the upper bound on η as we shall describe below. Deliyannis and collaborators have also discovered that an inclusion of

rotational effects which were not previously included in their codes can raise the upper bound on primordial Li considerably[23]. This effect could in principle alter the upper limit by an order of magnitude. However, since the effect of rotation is not yet clearly known, I present, merely for comparison purposes, bounds on η derivable by comparing the new BBN estimates with the Li observational limits quoted in ref 14. Using the ^{7}Li/H limits, one finds that the 2σ curve would yield an upper bound of $\eta_{10}=5.5$, down by about 15% from the previous estimate[14]. If one instead utilized an upper limit on primordial ^{7}Li/D, which is presumably more conservative because significant Li destruction in stars would be accompanied by even more D destruction, one would find an upper limit of $\eta_{10}=7.3$, a reduction of about 10% from that previously quoted. This would correspond to an upper limit on the fraction of the closure density in baryons of $\Omega_b h_o^2 < .028$.

However, a more stringent upper bound on η may be near, coming from the observational limits on ^{4}He. As we pointed out above, as the upper bound on ^{4}He is lowered, it becomes difficult to reconcile even three neutrinos with the standard BBN picture. For an upper limit of 24% by weight, one finds that the 2σ prediction crosses this value for $\eta_{10}=4.8$! (This upper limit is reduced to 4.5 if one were to utilize the new neutron lifetime measurement.) This bound would definitively establish that $\Omega_B < 0.08$ (for h$\geq$1/2), confirming the need for non baryonic dark matter to account for the dark halos in clusters. Because the neutron half-life is now measured to much less than 1%, the impact of such a comparison of theory with observation will not change significantly. It merely remains to improve observational limits on ^{4}He, in order to fix almost completely the baryon density today which is compatible with BBN. **Helium once again can become the great determinator in BBN.**

There is also another amusing alternative. It is quite possible that ^{7}Li might eventually restrict η_{10} to be greater than 3-4. If this is the case, and if the observational limit on ^{4}He continues to decrease, this would pose the first hint of a possible inconsistency in standard BBN---namely even three neutrinos might be too many! While this possibility is interesting, however, no such inconsistency is yet apparent. Raising the upper bound on ^{4}He to 24.5% raises the upper bound on η_{10} to about 8.5. Thus, one would have to be quite sure of one's upper limits on He before one would be able to claim that one of the known neutrinos is too heavy to be relativistic during BBN.

Finally, where can standard BBN go from here? Which experimental and observational improvements would be most significant? Clearly, as we have stressed above, if the observational uncertainties in ^{4}He can be reduced to the point where an upper or lower limit in the range of 24% could be definitively established, this would provide perhaps the strongest restriction on cosmology from BBN. Since the neutron half life uncertainty has been reduced, fixing the primordial abundance of He would essentially fix η. *It is also encouraging that were it not for the new lower limits on the neutron half life, BBN predictions would now be inconsistent for a ^{4}He abundance less than about 24%, for any value of $\eta_{10} > 2$-3. Whenever new experimental results improve agreement with theory,*

this bodes well for theory.

The chief remaining area where tangible improvements might quickly come about relates to the primordial abundance of ^{7}Li. Here, the limits can easily be improved if several key reactions are re-measured. Once again, by far the largest uncertainty in the use of ^{7}Li in comparing BBN predictions with observations lies in the poorly understood theoretical situation of Li evolution in stars. While recent work has demonstrated that no non-standard evolution is necessary to fit observations of type II stars, if the recent results indicating that rotation can lead in some sense to a uniform depletion of Li can be established quantitatively, then the lower limit on the ^{7}Li BBN abundance inferred from observation might be raised. In this case, it is possible that the lowest part of the predicted Li valley might fall outside of the allowed range based on observation. This would imply two disjoint regions of allowed η. Since the lower region is probably below that which may be ruled out by the ^{3}He+D limits, this could cause the new lower limit on η_{10} from ^{7}Li to raise up to almost 4.

Unfortunately, it is unlikely that any single experimental remeasurement will dramatically alter the present predictive power of the standard BBN scenario. It does seems likely that modest experimental improvments combined with improvements in observations, particularly of ^{4}He and ^{7}Li may, under optimistic circumstances, allow the range of η_{10} to be restricted to lie between about 4-5. *This would clearly demonstrate the existence of baryonic dark matter, but also the need for significant non-baryonic dark matter in the universe.*

BBN, in its simplest, homogeneous form, has provided a spectacularly successful model of the evolution of matter in the early universe. General agreement with the observed abundance of light elements has been obtained, and a prediction that the number of light neutrino types was less than 4-5 has been vindicated by the recent direct limits obtained at SLC and LEP. To go further in the attempt to place detailed limits on various cosmological and particle physics parameters it is necessary finally to take proper account of the uncertainties, not just in observations, but also in the experimental parameters which serve as input into BBN calculations. It is only in this way that BBN as a theory with predictive power can proceed from precocious youth to full maturity. I hope that the work described here is a step in this direction.

REFERENCES:
1. Applegate, J.H., C. J. Hogan, R.J.Scherrer 1987 *Phys. Rev.* **D35**, 1151
2. Fuller, G.M, G. Matthews, and C.R. Alcock 1987, *Ap. J.***320**, 439
3.Dimopoulos, S. , R. Ezmailzadeh, L.J. Hall, G. J. Starkman. 1989, *Ap. J.* **330**, 545
4. The results on BBN I report on here were obtained in collaboration with Paul Romanelli. These results (and many of the discussions) are taken verbatim from L. M. Krauss and P. Romanelli, Yale preprint (Oct 89) YCTP-P1-89, Astrophysical Journal, to appear July 20th 1990 issue. Further references to the material described here may be found there.
5.Wagoner, R.V. 1969, *Ap. J. Suppl.,* **18** (No. **162**), 247; 1973, *Ap. J.,* **179**, 343.
6. Kawano, L. 1988, **FERMILAB-PUB-88/34-A** Fermilab preprint.

122

7. i.e. see S. Weinberg, *Gravitation and Cosmology,*(Wiley, New York) 1972; Yang, J., M.S. Turner, G. Steigman, D.N. Schramm, and K. A. Olive 1984, *Ap. J.,* **281**, 493 (YTSSO); or E. W. Kolb and M. Turner, *The Early Universe,* (Benjamin Cummings, Menlo Park) 1990

8. i.e. see Deliyannis, C.P., P. Demarque, S.D.Kawaler, L.M.Krauss, P. Romanelli 1989, *Phys. Rev. Letters,* **62**, 1583; Audouze, J., J. Silk 1989, *Ap. J.* **342** L5; Deliyannis, C.P., P. Demarque, and S.D.Kawaler 1990, *Ap. J. Suppl.* (in press); C. P. Deliyannis, Yale Ph. D. Thesis 1990; Malaney, R.A., C.R. Alcock 1989, Livermore preprint, UCRL-101226

9. Kajino, T. and A. Arima 1984, *Phys. Rev. Lett.,* **52**, 739.

10. see Kajino, T., H. Toki and S.M. Austin 1987, *Ap. J.,* **319**, 531 for a review

11. see ref. 9

12. Walliser, H., H. Kanada, and Y.C. Tang 1984, *Nucl. Phys.,* **A419**, 13

13. Rolfs, C., and R.W. Kavanaugh 1986, *Nucl. Phys.,* **A455**, 179 (RK)

14. Deliyannis, C.P., P. Demarque, S.D.Kawaler, L.M.Krauss, P. Romanelli 1989, *Phys. Rev. Letters,* **62**, 1583

15. Bahcall, N. A., and W.A. Fowler 1969, *Ap. J.,* **157**, 659

16. Koehler, P.E., C.D. Bowman, F.J. Steinkruger, D.C. Moody, G.M. Hale, S.W. Starner, S.A. Wender, R.C. Haight, P.W. Lisowski, and W.L. Talbert 1988, *Phys. Rev. C,* **37**, 917

17. Particle Data Group 1988, *Phys. Letters,* **B204**, 1

18. Mampe, W., P. Ageron, C. Bates, J.M. Pendlebury, and A. Steyerl 1989, *Phys. Rev. Letters,* **63**, 593

19. Bahcall, J.N., W.F. Huebner, S.H. Lubow, P.D. Parker, and R.K. Ulrich 1982, *Rev. Mod. Phys.,* **54**, 767

20. Yang, J., M. Turner, G. Steigman, D. Schramm, and K. Olive 1984, *Ap. J.,* **281**, 493

21. L3 , *Phys. Lett* **B231**, 509 (1989); ALEPH, *Phys. Lett* **B231**, 519 (1989) ; OPAL, *Phys. Lett* **B231**, 530 (1989); Delphi, *Phys. Lett* **B231**, 539 (1989); G.S Abrams et al. *Phys. Rev. Lett.* **63**, 2173 (1989)

22. L. M. Krauss, *Phys. Rev. Lett.* **64**, 999 (1990)

23. Pinsonneault, M.H., C.P. Deliyannis, and P. Demarque 1989, in preparation; C.P. Deliyannis, Yale Ph.D. Thesis 1990

Charged Bosonic Stellar Configurations

Philippe Jetzer

CERN, CH-1211 Geneva 23, Switzerland

Abstract

I present the static spherically symmetric gravitational equilibria of scalar fields coupled to a U(1) gauge field and with a possible $\frac{\lambda}{4}(\phi^*\phi)^2$ self-interaction. These configurations are obtained by solving numerically the coupled system of Einstein-Maxwell-Klein-Gordon equations for non-singular and asymptotically flat solutions. Static solutions only exist for values of the gauge coupling constant such that $e^2/4\pi \leq G_N m^2$, where m is the mass of the scalar particle. The maximum mass of the Bose star increases with increasing value of the gauge coupling constant. I discuss also the dynamical stability of the equilibrium configurations, for which I derive a pulsation equation, which governs the time evolution of the infinitesimal radial oscillations, as well as a variational principle for its eigenvalues. The equilibrium configurations with a central density bigger than ρ_{crit}, corresponding to the critical mass, are dynamically unstable.

1 Introduction

The more recent developments in particle physics and cosmology suggest that scalar fields may have played an important role in the evolution of the early universe, for instance in primordial phase transitions and that they may make up part of the missing dark matter. Models for galaxy formation using cold dark matter and the inflationary scenario suggest that the ratio of baryonic (luminous) matter to dark matter can be of the order of 10 %. These facts naturally raise the question whether cold gravitational equilibrium configurations of massive scalar fields - Bose stars - may exist and whether such configurations are dynamically stable. Spherically symmetric equilibrium configurations of scalar fields have been found by solving the coupled Einstein-Klein-Gordon equations without [1-4] or with a quartic self-interaction [5]. It is also possible to consider other types of objects made up of scalar fields by the inclusion of gravity in non-topological solitons [6] and in Q-balls [7]. In all these models, if one plots the mass for equilibrium configurations of Bose stars against their central density, one finds a behaviour very similar to that of neutron stars. The mass quickly rises to a maximum for $\rho = \rho_{crit}$, drops a little, oscillates and approaches an asymptotic value at large central densities. A similar behaviour is found for the particle number.

Here I discuss the static spherically symmetric solutions for a system of complex scalar fields coupled to a U(1) gauge field [8]. I also allow for the possibility of a quartic self-interaction. The main effect of having scalar fields coupled to a U(1) gauge field is to increase the maximum mass with increasing gauge coupling constant e. At the same time the binding energy per particle decreases. Static solutions only exist for $e^2/4\pi \leq G_N m^2 = e_{crit}^2/4\pi$. I consider also the problem of the dynamical stability of the equilibrium configurations in the framework of general relativity [9]. I analyse the time evolution of infinitesimal radial oscillations, which conserve the total number of particles, following the method developed by Chandrasekhar [10]. I find an eigenvalue equation, called also pulsation equation, which determines the normal modes of the radial oscillations and a variational principle for determining the eigenvalues [9,11]. As in the cases without charge [4], the particle number and mass have their extrema , in particular their maximum, at the same value of the central density. From this fact it follows that the pulsation equation has a zero mode, where M and N have their extrema [12,13]. Therefore one expects the central density ρ_{crit}, corresponding to the maximum mass, to be the boundary between stable (for ρ smaller than ρ_{crit}) and unstable equilibrium configurations. However in order to establish this fact completely one has to analyse the eigenvalues of the pulsation equation to see if they are real and positive for central densities smaller than ρ_{crit}. M and N as a function of the central density have other stationary points for ρ bigger than ρ_{crit}. We expect the behaviour of the stability not to change there, a fact which is already

suggested by the shape of the particle number versus radius diagram [8], which is bent counterclockwise at the critical points. A rigorous proof, however, can be given using the variational principle; in fact one can easily find an upper bound to the lowest eigenvalue at the stationary points of M and N for ρ bigger than ρ_{crit}. It turns out that the upper bounds are negative, therefore the lowest eigenvalue is negative which establishes that the equilibrium configurations remain unstable for all ρ bigger than ρ_{crit}.

2 Equilibrium configurations

The system we consider is a complex scalar field with a U(1) charge coupled to gravity, whose action is given by

$$S = \int d^4x \sqrt{-g} \left(\frac{R}{16\pi G_N} + g^{\mu\nu}(D_\mu \phi)^*(D_\nu \phi) - m^2|\phi|^2 - \frac{\bar{\lambda}}{2}|\phi|^4 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \right), \qquad (1)$$

with $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ and $D_\mu \phi = \partial_\mu \phi + ieA_\mu \phi$. The action is invariant under a local U(1) gauge transformation, thus the total particle number is conserved. We take all the fields as functions of r and t alone since we consider only spherically symmetric equilibrium configurations. We express the metric in Schwarzschild coordinates

$$ds^2 = e^\nu dt^2 - e^\lambda dr^2 - r^2(d\vartheta^2 + sin^2\vartheta d\varphi^2) \qquad (2)$$

where ν and λ are functions of r and t only ($g^{00} = -e^{-\nu}$ and $g^{rr} = e^{-\lambda}$). We choose the gauge field A_μ such that we have only electric charges and no magnetic ones, therefore we set $A_\mu = (A_t = A, A_r = C = 0, A_\vartheta = 0, A_\varphi = 0)$. We write the complex scalar field as $\phi = (\phi_1 + i\phi_2)e^{i\omega t}$, where ϕ_1 and ϕ_2 are two real fields. For the equilibrium configuration we set $\phi_1(r, t) = \phi_0(r)$, $\phi_2(r, t) = 0$ and all the other fields and metric functions are time independent. The coupled system of equations, which can be derived from the above action by varying with respect to the various fields, is given by the two Einstein equations

$$\lambda_0' = \frac{(1 - e^{\lambda_0})}{r} + 8\pi G_N r e^{\lambda_0}[((\omega + eA_0)^2 e^{-\nu_0} + m^2)\phi_0^2 + \frac{\bar{\lambda}}{2}\phi_0^4 + \phi_0'^2 e^{-\lambda_0} + \frac{A_0'^2}{2}e^{-\nu_0 - \lambda_0}], \quad (3)$$

$$\nu_0' = \frac{(e^{\lambda_0} - 1)}{r} + 8\pi G_N r e^{\lambda_0}[((\omega + eA_0)^2 e^{-\nu_0} - m^2)\phi_0^2 - \frac{\bar{\lambda}}{2}\phi_0^4 + e^{-\lambda_0}\phi_0'^2 - \frac{A_0'^2}{2}e^{-\nu_0 - \lambda_0}], \quad (4)$$

the Maxwell equation

$$A_0'' + (\frac{2}{r} - \frac{\nu_0' + \lambda_0'}{2})A_0' - 2e\phi_0^2 e^{\lambda_0}(\omega + eA_0) = 0, \qquad (5)$$

and the scalar wave equation

$$\phi_0'' + (\frac{2}{r} + \frac{\nu_0' - \lambda_0'}{2})\phi_0' + e^{\lambda_0}((\omega + eA_0)^2 e^{-\nu_0} - m^2 - \bar{\lambda}\phi_0^2)\phi_0 = 0. \qquad (6)$$

As boundary conditions we impose $\phi_0(0)$=const, $\phi'_0(0)=0$ and $\phi_0(\infty)=\phi'_0(\infty)=0$ in order to have a localized particle distribution. We allow the possibility of ϕ_0 having nodes, giving rise to excited states. Furthermore we demand that the electric field be absent at the origin, giving $A'_0(0)=0$, and we normalize the potential by $A_0(\infty)=0$. The condition $e^{\nu_0(\infty)}=1$ is imposed to get asymptotically the ordinary Minkowski metric and the condition $e^{\lambda_0(0)}=1$ is a regularity condition. With these boundary conditions the equations (3-6) form an eigenvalue equation for ω. For a detailed discussion of the equilibrium solutions see Ref.[8].

The total mass of the system is determined by the integral

$$M = 4\pi \int_0^\infty dr r^2 [((\omega + eA_0)^2 e^{-\nu_0} + m^2)\phi_0^2 + \frac{\tilde{\lambda}}{2}\phi_0^4 + e^{-\lambda_0}\phi_0'^2 + \frac{A_0'^2}{2}e^{-\nu_0 - \lambda_0}], \qquad (7)$$

and the total particle number N_0, which is related to the total charge by $Q = eN_0$, is given by

$$N_0 = 8\pi \int_0^\infty dr r^2 (\omega + eA_0)\phi_0^2 e^{(\nu_0 - \lambda_0)/2}. \qquad (8)$$

The effective radius of the Bose star can be defined as

$$R = 8\pi \int_0^\infty dr r^3 (\omega + eA_0)\phi_0^2 e^{(\lambda_0 - \nu_0)/2}. \qquad (9)$$

Some results of the numerical solutions of the equations (3)-(6) are shown in Fig. 1-3. For values of the charge close to the critical critical charge $\tilde{e}_{crit}=\sqrt{1/2}$ ($\tilde{e}_{crit}$ and $\tilde{e}$ are in units of $M_{Pl}/\sqrt{8\pi}m$) the maximum mass has the following asymptotic bahaviour

$$M_{max} \approx 0.44(\tilde{e}_{crit} - \tilde{e})^{-1/2}\frac{M_{Pl}^2}{m} \qquad (10)$$

for $\tilde{\lambda} = 0$ and

$$M_{max} \approx 0.226(\tilde{e}_{crit} - \tilde{e})^{-1/2}\frac{M_{Pl}^3}{m^2}\sqrt{\frac{\tilde{\lambda}}{8\pi}}. \qquad (11)$$

for $\tilde{\lambda} \neq 0$. Similarly for the value of the radius corresponding to the maximum mass we get

$$R \approx 1.5(\tilde{e}_{crit} - \tilde{e})^{-1/2}\frac{1}{m} \qquad (12)$$

for $\tilde{\lambda} = 0$ and

$$R \approx 0.415(\tilde{e}_{crit} - \tilde{e})^{-1/2}\frac{M_{Pl}}{m^2}\sqrt{\frac{\tilde{\lambda}}{8\pi}} \qquad (13)$$

for $\tilde{\lambda} \neq 0$.

Table 1: Comparison between different types of Bose stars

	Mass	Radius
$\tilde{\lambda}=0$, $\tilde{e}=0$	$6.1\ 10^{-19}\ M_s$	$4.2\ 10^{-15}$
$\tilde{\lambda}=0.1$, $\tilde{e}=0$	$1.2\ M_s$	$5.8\ 10^3$
$\tilde{\lambda}=\bullet$, $\tilde{e}=0.5$	$9.2\ 10^{-19}\ M_s$	$5.6\ 10^{-15}$
$\tilde{\lambda}=0.1$, $\tilde{e}=0.5$	$2.6\ M_s$	$8.5\ 10^3$

The values in the table are for a boson mass m=$m_\pi \approx 139$ MeV and M_s=2 10^{30} kg is the solar mass. The values for the radius are in metres. We notice that by suitably choosing the parameters $\tilde{\lambda}$, $\tilde{e}$ and the boson mass m one can construct Bose stars, which can support very fast rotations without breaking up [14]. This because, as can be seen from the values of the above table by comparing for instance the $\tilde{\lambda} \neq 0$ case with models for neutron stars with comparable masses, the Bose stars can be very centrally condensed.

3 Pulsation equation and variational principle

Consider now the situation where the equilibrium configuration is perturbed in a way such that the spherical symmetry is still preserved. These perturbations will give rise to motions in the radial direction. The equations governing the small perturbations are obtained by expanding all functions to first order and by linearizing the equations. By suitably combining the perturbation equations [9,11] and supposing a time-dependence of the form $e^{i\sigma t}$, one gets the following eigenvalue equation for f_1 and f_2, which are related to the radial part of $\delta\phi_1(r,t), \delta\phi_2(r,t)$ and $\delta C(r,t)$ [9]

$$L_{ij}f_j = \sigma^2 M_{ij}f_j \qquad i,j = 1,2 \tag{14}$$

where

$$M_{ij} = e^{\lambda_0 - \nu_0}\begin{pmatrix} G_1 & 0 \\ 0 & G_2 \end{pmatrix}$$

with $G_1 = r^2 e^{(\nu_0 - \lambda_0)/2}$ and $G_2 = r^2 \phi_0^2 e^{(\nu_0 - \lambda_0)3/2}$,

$$L_{ij} = \begin{pmatrix} -\frac{\partial}{\partial r}G_1\frac{\partial}{\partial r} + G_1 C_1 & -2\frac{\partial}{\partial r}((\omega + eA_0)G_1\phi_0) + G_1 C_3 \\ 2(\omega + eA_0)G_1\phi_0\frac{\partial}{\partial r} + G_1 C_3 & -\frac{\partial}{\partial r}G_2\frac{\partial}{\partial r} + G_2 C_2 \end{pmatrix}$$

with

$$C_1(r) = e^{\lambda_0}(3e^{-\nu_0}(\omega + eA_0)^2 + m^2 + \frac{\tilde{\lambda}}{2}3\phi_0^2) - 8\pi G_N r\phi_0'^2(\frac{2}{r} + \nu_0' - \lambda_0') \tag{15}$$
$$+ 32\pi G_N r e^{\lambda_0}\phi_0'(m^2\phi_0 + \frac{\tilde{\lambda}}{2}\phi_0^3),$$

$$C_2(r) = \frac{2(\lambda_0' - \nu_0')}{r} + \frac{2}{r^2} - \frac{\nu_0'' - \lambda_0''}{2} + 2(\frac{\phi_0'}{\phi_0})^2 - 2\frac{\phi_0''}{\phi_0} + \frac{(\lambda_0' - \nu_0')}{2}(\nu_0' - \lambda_0' + 4\frac{\phi_0'}{\phi_0}) + 2e^2 e^{\lambda_0}\phi_0^2$$
$$- 8\pi G_N r e^{\lambda_0 - \nu_0}(\omega + eA_0)^2\phi_0^2(\nu_0' - \lambda_0' + \frac{2}{r}) + 32\pi G_N r(\omega + eA_0)\phi_0^2 e^{\lambda_0 - \nu_0} eA_0', \tag{16}$$

$$C_3(r) = -2\phi_0'(\omega + eA_0) - 8\pi G_N r(\omega + eA_0)\phi_0'(\nu_0' - \lambda_0' + \frac{2}{r})\phi_0^2 + 2e\phi_0 A_0' \tag{17}$$
$$+ 16\pi G_N r e^{\lambda_0}(\omega + eA_0)(m^2\phi_0 + \frac{\tilde{\lambda}}{2}\phi_0^3)\phi_0^2 + 16\pi G_N r e A_0'\phi_0'\phi_0^2.$$

Eq.(14) is the required "pulsation equation". The appropriate boundary conditions are for $r \to \infty$: $f_1 \to 0$, $r^2 \phi_0^2 f_2 \to 0$ and for $r \to 0$: f_1=const, $r^2 f_2 \to 0$. With these boundary conditions both the operators L_{ij} and M_{ij} are symmetric, this also in the case where ϕ_0 has nodes [9,15], and the total particle number is automatically a conserved quantity [9]. The eigenvalues are real and dynamical instability will occur whenever $\sigma_0^2 \leq 0$.

Eq.(14) can also be obtained from the following variational principle

$$\sigma^2 \int_0^\infty \frac{1}{2} e^{\lambda_0 - \nu_0} (G_1 f_1^2 + G_2 f_2^2) dr = \int_0^\infty [\frac{1}{2} G_1 f_1'^2 + \frac{1}{2} G_1 f_1^2 C_1 + \frac{1}{2} G_2 f_2'^2 + \frac{1}{2} G_2 f_2^2 C_2 \\ + f_2 f_1' 2 \phi_0 (\omega + \epsilon A_0) G_1 + f_2 f_1 G_1 C_3] dr. \tag{18}$$

A sufficient condition for dynamical instability to occur is that the right-hand side of eq.(15) vanishes (or becomes negative) for some chosen pair of trial functions f_1 and f_2 which satisfy the above boundary conditions. It turns out that for central densities bigger than ρ_{crit} the equilibrium configurations are unstable.

References

[1]   D.J. Kaup, Phys. Rev. 172 (1968) 1331.

[2]   R. Ruffini and S. Bonazzola, Phys. Rev. 187 (1969) 1767.

[3]   J.D. Breit, S. Gupta and A. Zaks, Phys. Lett. 140B (1984) 329.

[4]   T.D. Lee, Phys. Rev. D35 (1987) 3637; R. Friedberg, T.D. Lee and Y. Pang, Phys. Rev. D35 (1987) 3640.

[5]   M.Colpi, S.L. Shapiro and I. Wasserman, Phys. Rev. Lett. 57 (1986) 2485.

[6]   R. Friedberg, T.D. Lee and Y. Pang, Phys. Rev. D35 (1987) 3658.

[7]   B.W. Lynn, Nucl. Phys. B321 (1989) 465.

[8]   Ph. Jetzer and J.J. van der Bij, Phys. Lett. B227 (1989) 341.

[9]   Ph. Jetzer, Phys. Lett. B231 (1989) 433.

[10] S. Chandrasekhar, Phys. Rev. Lett. 12 (1964) 114,437; Astrophys. J. 140 (1964) 417.

[11] Ph. Jetzer, Nucl. Phys. B316 (1989) 411; Proc. Fifth Marcel Grossmann Meeting (Perth,August 1988) (World Scientific, Singapore) p. 1249.

[12] T.D. Lee and Y. Pang, Nucl. Phys. B315 (1989) 477.

[13] Ph. Jetzer, CERN-preprint TH 5685 (1990) to appear in Phys. Lett. B.

[14]   S. Shapiro, S. Teukolsky and I. Wasserman, Nature 340 (1989) 451.

[15]   Ph. Jetzer, Phys. Lett. B222 (1989) 447; CERN-preprint TH 5454 (1989),in: Proc.
       First Topical Seminar on Astrophysics and Particle Physics (San Miniato, May 1989),
       to appear.
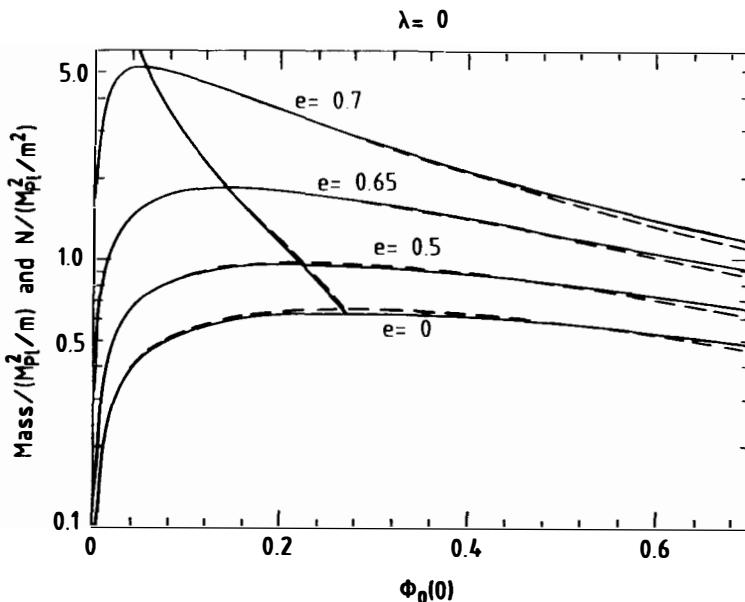
Fig. 1:   Boson star mass in units of M_{Pl}^2/m (solid line) and particle number in units of
M_{Pl}^2/m^2 (broken line) as a function of $\phi_0(0)$. The charge e is given in units of
$M_{Pl}/(\sqrt{8\pi}m)$. No quartic self-coupling is present. Also the lines going through
the maxima of mass and particle number are drawn.
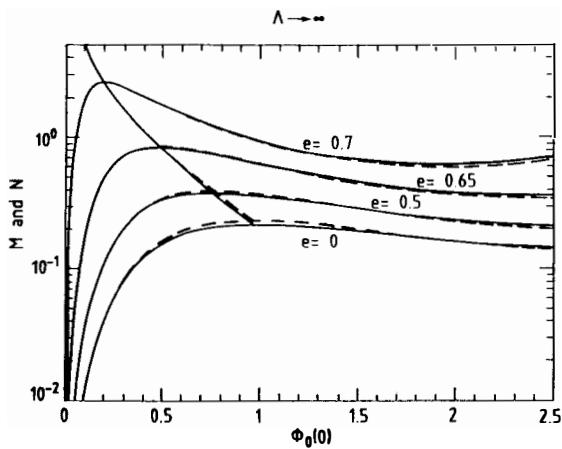
Fig. 2: Boson star mass in units of $\Lambda^{1/2}M_{Pl}^2/m$ (solid line) and particle number in units of $\Lambda^{1/2}M_{Pl}^2/m^2$ (broken line) as a function of $\phi_0(0)\Lambda^{1/2}$ for different values of e (in units of $M_{Pl}/\sqrt{8\pi}m$) for the case $\Lambda \to \infty$. Also the lines going through the maxima of mass and particle number are drawn.
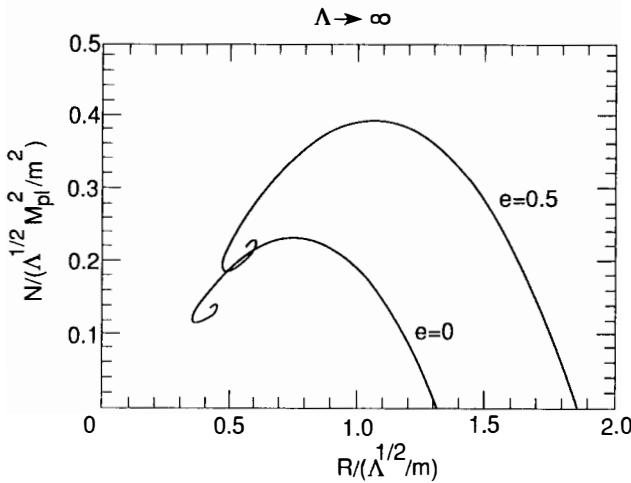
Fig. 3: The particle number in units of $\Lambda^{1/2}M_{Pl}^2/m^2$ as a function of the effective radius R (in units of $\Lambda^{1/2}/m$) for the $\lambda \neq 0$ case for two different values of the charge e (in units of $M_{Pl}/\sqrt{8\pi}m$).

BARYON Q-BALLS: A NEW FORM OF MATTER?

Stephen B. Selipsky

Department of Physics
Stanford University, Stanford, California 94305

Abstract

Hadronic effective field theories describing ordinary nuclei also contain Q-Ball solutions which can describe a new state of matter, "baryon matter". Baryon matter is stable in very small chunks as well as in stellar-sized objects, since it is held together by the strong force instead of just gravity. Larger chunks, "Q-Stars", in which gravity is important, model neutron stars. A wide variety of Q-Star models, all consistent with known nuclear physics, allow compact objects to have masses much larger, or rotation periods much shorter, than is conventionally believed possible. Smaller chunks of nuclear density baryon matter could also be astrophysically important components of the universe, and at late times would have many properties similar to those of strange matter chunks.

1. Introduction

There are many surprising possibilities lurking in the non- perturbative sector of field theories. Here we add to the body of work on this topic, presenting the possibility of a new state of matter which arises from the discovery of solutions to effective field theories describing among other things ordinary nuclei. Effective field theories of interacting baryons and mesons have successfully reproduced measured properties of nuclei as well as results of scattering experiments[1-3]). We have found[4-6]) non-topological classical solutions to such theories: fermion Q-Balls, or Q-Stars in the case where their self–gravity is important and gravitational effects are included. Q here stands for the conserved charge (baryon number) which stabilizes the matter against decay. The properties of such a state of baryon matter can be different from and essentially independent of the characteristics of ordinary nuclei studied in the laboratory. In particular, neutron stars may have a large binding energy per nucleon, hundreds of MeV, due mainly to nuclear forces, and may be more massive or able to rotate faster than is suggested by currently accepted limits.[7,8]) In addition, chunks of baryon matter, varying in size from 10^{-12}cm to several kilometers, may also exist. Structural characteristics are fairly insensitive to the effective field theory used; classes of theories give the same equation of state.

2. Constructing Solutions

There is a simple graphical method for finding large baryon-number solutions. Consider a Lagrangian for a baryon ψ interacting with a scalar and vector field:

$$\mathcal{L} = \bar{\psi}\left[i\partial\!\!\!/ - m(\sigma) - g_v V\!\!\!\!/\right]\psi + \tfrac{1}{2}(\partial_\mu \sigma)^2 - U(\sigma) + \tfrac{1}{2}m_v^2 V_\mu V^\mu . \tag{2.1}$$

It is a good approximation to neglect the dynamics of the vector field[1]), and for a many-fermion system we can make the Thomas-Fermi approximation ($dm/dr \ll m^2$) to get a Fermi sea of baryons described by a constant chemical potential ε_F and a Fermi momentum k_F slowly varying in space. The identity $\langle \bar{\psi}\psi \rangle = -(\partial P_\psi/\partial m)^{6})$ gives the classical

equation of motion

$$\nabla^2\sigma \;=\; -\frac{\partial}{\partial\sigma}(P_\psi - U) \tag{2.2}$$

where P_ψ is the pressure of the baryons. This is equivalent to Newton's $F = ma$ for a mechanical 'particle' at 'position' σ at 'time' r moving in a potential $V_{eff} = P_\psi - U$, with 'friction' from the $(2/r)\partial\sigma/\partial r$ term negligible for large r.

The Q-Ball solution occurs when σ rolls between degenerate maxima of V_{eff}, where one of the maxima is the vacuum and the height of the other is tuned by varying ε_F. The σ field starts at a value σ_{inside} at $r = 0$, close to the top of the first hill, and stays near that value out to a large radius (since the top of the hill is flat) before quickly rolling to the vacuum. The large baryon number Q-Ball thus has a flat interior, a radius which is a free parameter, and a thin surface (of order the scalar Compton wavelength). The density is determined by algebraic equations for ε_F and σ_{inside}, requiring degenerate maxima: $V_{eff} = 0$ and $\partial V_{eff}/\partial\sigma = 0$. A Q-Ball must still have total energy $E < Qm_N$ in order to be bound. Further details can be found in Refs. 6 and 9.

3. Application

We can apply this construction quite generally, for instance to Walecka's Quantum Hadro-Dynamics[1]), in which the proton and neutron are fermion fields ψ_i whose effective mass is $m(\sigma) = g_s\sigma$, with the free–particle $m_N = g_s\sigma_0$. The scalar field has a quadratic potential $U = \frac{1}{2}m_s^2(\sigma - \sigma_0)^2$, and the vector field is the ω particle. This theory's large baryon number Q-Ball corresponds to infinite symmetric nuclear matter when $N = Z$. The small baryon number solutions (nuclei) depend on the friction term in eq. (2.2); the rolling starts above $P_\psi - U = 0$ and friction brings σ to rest at the vacuum value where $m = m_N$. Fig. 1 translates the Walecka solutions into this framework.

Infinite nuclear matter is destabilized by Coulomb forces. In fact, neutrons, protons, and electrons coexist in β-decay equilibrium, allowing local charge neutrality: $k_{F,p} = k_{F,e}$
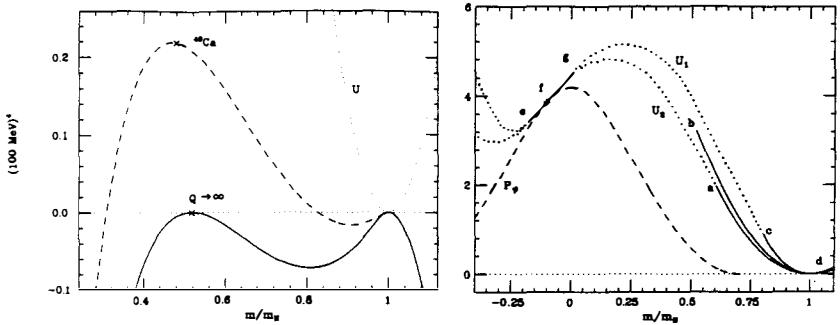
Figure 1. Graphical representation of Q-Balls in a nuclear theory. $U(\sigma)$ (dots) is the scalar field potential, P_ψ is the baryon pressure, and $P_\psi - U$ is the effective potential that σ 'rolls' in, starting at $r = 0$ (marked by crosses).

Figure 2. Potentials U (solid lines) given in Refs. 1 and 2.

and $\varepsilon_{F,n} = \varepsilon_{F,p} + \varepsilon_{F,e}$. With a baryon pressure function reflecting this, a large baryon number Q-Ball does not exist in Walecka's theory, so strong gravity for stellar sized objects is the only possibility. However we are still free to choose other potentials. In an effective field theory the potential and the coupling $m(\sigma)$ are in general nonrenormalizable, subject only to the symmetry of the underlying theory, and should either be fit from experiment (à la Walecka) or derived from the underlying theory (QCD). Then m and U have quantum corrections included in their definitions. Boguta, Strocker, and others have found many models with renormalizable potentials which reproduce known bulk properties of nuclear matter[2]). Fig. 2 shows these and Walecka's potentials; the region 'a-b-c-d' is the only part relevant for nuclear data, and U is not constrained elsewhere, allowing infinitely many altered potentials (*e.g.* dotted lines) that do give large baryon number Q-Balls. Only the (currently unmeasurable) point 'f' determines Q-Ball bulk properties, and the rest of U affects only the thin surface. In this way we can have stable macroscopic chunks of high density baryon matter, of any size above Q_{bulk} where surface effects become important. To fix point 'f' and the properties of neutron stars, experiments on baryon matter chunks would be necessary.

The simple 'chiral' Q-Ball case in which the bulk-phase nucleon mass $m(\sigma_{inside}) = 0$, has features generic to more complicated models (which have also been solved). The equation of state[9]) for a chiral Q-Star (neglecting the electron mass and neutron-proton mass difference) depends only on the vector repulsion strength and $U_0 \equiv U(\sigma_{inside})$:

$$\mathcal{E} - 3P - 4U_0 + \alpha_{\mathbf{v}}(\mathcal{E} - P - 2U_0)^{3/2} = 0 \ , \qquad (3.1)$$

with $\alpha_{\mathbf{v}} \equiv (g_{\mathbf{v}}/m_{\mathbf{v}})^2 3^{1/2}/\pi$. Only the nucleons couple to the strong forces, so charge separation results in the surface. Structure calculations may neglect these details since the surface width is about 10^{-20} times smaller than the stellar radius, but the electrostatic gap means that small Q-Balls will not absorb ordinary matter and Q-Stars can support a crust of degenerate ordinary matter below neutron drip densities.

Since Q-Balls are a bulk phase with an equation of state, gravity can be included simply by integrating the Oppenheimer-Volkoff equations. On dimensional grounds, when $GM/R \sim 1$ a Q-Star has radius $R \sim m_{pl}\sigma_0^{-2} \sim 100$ km and mass $M \sim m_{pl}^3\sigma_0^{-2} \sim 10 \ M_\odot$, where $\sigma_0 \sim 100 MeV$ is the vacuum value of σ. Integrating the Oppenheimer-Volkoff equations for eq. (3.1) gives the solid lines in Fig. 3. We also show for comparison (dashed lines) two neutron star models (pion condensation and Walecka's) which are respectively among the least stiff and most stiff of conventional neutron star models. In conventional models the stellar radius decreases as the mass and baryon number increase, while Q-Balls have M and $Q \propto R^3$, giving the generic curlique.

Both stellar radius and mass increase as the unknown U_0 decreases, allowing a large upper limit for the neutron star mass if U_0 is small. However eq. (3.1) is only valid at densities high enough for the nuclear interactions to dominate. There seem to be no experimental constraints on the equation of state of a *large number* of baryons above white dwarf densities ($< 10^{10}$gm/cm^3), which we will take as the lowest permissible U_0 value. Using (3.1) down to density U_0 evades the theoretical Rhoades-Ruffini limit, $3.2 M_\odot$, which assumes some conventional equation of state below a density $\mathcal{E}_0$ taken to be about
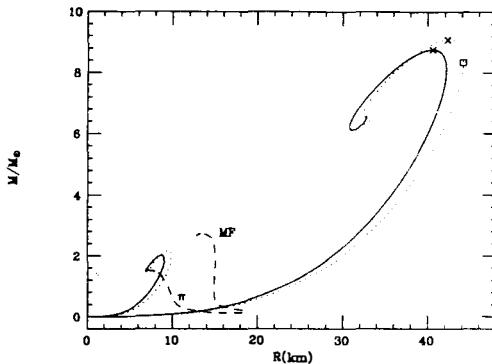
Figure 3. Mass $vs.$ stellar radius for chiral (solid lines) and representative non-chiral (dotted lines) Q-Stars, with $U_0 = (85MeV)^4$ (small curves) and $(200MeV)^4$ (large). Central energy density increases along the curliques; the maxima (crosses) are the last stable configurations.

nuclear density.

For $\alpha_v U_0^{1/2} = 1.23$ (Walecka's value when $U_0 = (100MeV)^4$), the upper mass limit on a neutron star with equation of state (3.1) is

$$M_{\max} = 6.68 \, U_{100}^{-1/2} \, M_\odot \, , \tag{3.2}$$

where $U_{100} \equiv U_0/(100MeV)^4$. The star also reaches its maximum baryon number at this mass: $Q_{max} = 22.5 \, U_{100}^{-3/4} \times 10^{57}$. The scaling with U_0, including $R \sim U_0^{-1/2}$, holds only if α_v is simultaneously scaled as $U_0^{-1/2}$; for bound configurations without this scaling,

$$0.48 M_\odot \; < \; M_{\max} \; < \; 400 M_\odot; \;\; 2.8 \text{ km} \; < \; R_{\max} < \; 2300 \text{ km}$$

$$0.67 \times 10^{57} \; < \; Q_{\max} < \; 1.7 \times 10^{60}; \; 3.4 \times 10^{43} \text{ g cm}^2 \; < \; I_{\max} < \; 2.0 \times 10^{52} \text{ g cm}^2 \tag{3.3}$$

Even small Q-Balls are stable against dispersal into free particles, because of their strong–interaction binding energy, and against adiabatic radial pulsations, since their adiabatic index $\Gamma \sim U_0(P_\psi + P_e - U)^{-1} \gg 4/3$. The turning point for the stability of this mode occurs at the point where $dM/d\mathcal{E}_{central} = 0$, indicated by crosses in Fig. 3, so Q-Stars before the maximum mass in Fig. 3 are stable. For a rotating Q-Star, the relativistic Maclaurin spheroid analysis as applied[8]) to other neutron star models shows

that dense Q-Stars can rotate at extremely high rates[9]), easily above the 0.5 ms limits on other equations of state. At the high densities needed, an effective field theory of baryons and mesons might not be applicable, but quark Q-Star models might apply. The chiral Q-Ball equation of state (neglecting vector repulsion) has the form of the MIT bag model's strange stars and nuggets[10]) (neglecting gluon exchange). The potential $U(\sigma)$ *dynamically* generates the analogue of bag pressure, but the magnitude of U_0 is of course unrelated to the nucleon bag pressure.

The Q-Star model differs in important respects from conventional neutron star models, raising interesting astrophysical questions. In addition to the stellar structure limits discussed above, electromagnetic properties of Q-Stars might be very different from those of neutron stars in conventional models. The neutrino cooling rate should also be much higher, since there is a boson condensate to absorb momentum for the first-order URCA process. Only a restricted class of effective field theories in which U wiggles such that there are two separate Q-Star phases (with different chemical potentials) can describe both neutron stars with large masses ($\gg 3M_\odot$) and neutron stars with short rotation periods ($P_{rot} < 0.5$ms). A conventional neutron star phase is always also present in any field theory which contains Q-Stars, and which phase is preferred will depend on the theory. In the early universe, a hadronic phase transition may well produce Q-Balls[11]), which (unlike strange nuggets) can be deeply enough bound to survive evaporation for $Q \gtrsim 10^{35}$. Q-Balls absorb neutrons and inhibit nucleosynthesis.[#1] In the present universe, baryonic dark matter can reside in remaining Q-Balls; many strange matter calculations still apply to the observability of a cosmic or galactic flux[12]), since these are also approximately nuclear-density lumps with a Coulomb barrier to fusion. For the larger Q-Balls that survive evaporation, direct detection would be difficult. However Q-Balls have quite a low surface tension and tend to break up into small droplets in any violent event. Neutron star collisions or possibly supernova explosions could produce some galactic flux of smaller

---

#1 I thank Jes Madsen for discussions on these points.

Q-Balls.

REFERENCES

1. B.D. Serot and J.D. Walecka in: Advances in Nuclear Physics, vol. 16, ed. J.W. Negele and E. Vogt (Plenum, New York, 1985), p. 1;
   B.D. Serot and J.D. Walecka, Phys. Lett. B 87 (1979) 172;
   J.D. Walecka, Annals Phys. 83 (1974) 491;
   F.E. Serr and J.D. Walecka, Phys. Lett. B 79 (1978) 10.

2. J. Boguta, Phys. Lett. B 106 (1981) 241;
   J. Boguta and S.A. Moszkowski, Nucl. Phys. A 403 (1983) 445;
   J. Boguta and H. Strocker, Phys. Lett. B 120 (1983) 289.

3. H. Georgi, Weak Interactions and Modern Particle Theory (Benjamin-Cummings, Menlo Park, CA, 1984).

4. For the theory of boson Q-Balls see
   R. Friedberg, T.D. Lee and A. Sirlin, Phys. Rev. D 13 (1976) 2739;
   S. Coleman, Nucl. Phys. B 262 (1985) 263.

5. The original Q-Star ansatz is given in B.W. Lynn, Nucl. Phys. B 321 (1989) 465;
   see also S.B. Selipsky, D.C. Kennedy, and B.W. Lynn, Nucl. Phys. B 321 (1989) 430.

6. For the theory of fermion Q-Balls and Q-Stars see S. Bahcall, B.W. Lynn, and S.B. Selipsky, Nucl. Phys. B 325 (1989) 606.

7. C.E. Rhoades and R. Ruffini, Phys. Rev. Lett. 32 (1974) 324;
   J.B. Hartle, Phys. Reports 46, No. 6 (1978) 201.

8. S. Shapiro, S.A. Teukolsky and I. Wasserman, Ap. J. 272 (1983) 702; Cornell preprint CRSR-919, May 1989;
   J.L. Friedman, J.R. Ipser and L. Parker, Nature 312 (1984) 255.

9. S. Bahcall, B.W. Lynn, and S.B. Selipsky, Nucl. Phys. B 331 (1990) 67;
   S. Bahcall, B.W. Lynn, and S.B. Selipsky, Astrophys. J., in press.

10. E. Witten, Phys. Rev. D 30 (1984) 272;
    E. Farhi and R. L. Jaffe, Phys. Rev. D 30 (1984) 2379;
    C. Alcock, E. Farhi and A. Olinto, Ap. J. 310 (1986) 261.

11. A.E. Nelson, Phys. Lett. B 240 (1990) 179.

12. A. de Rujula and S.L. Glashow, Nature 312 (1984) 734;
    A. de Rujula, Nucl. Phys. A 434 (1985) 605;
    P.B. Price, Phys. Rev. D 38 (1988) 3813.

DARK MATTER IN HALOES

OBSERVATIONAL CONSTRAINTS ON BARYONIC DARK MATTER

Bernard Carr
School of Mathematical Sciences,
Queen Mary & Westfield College,
Mile End Road,
London E1 4NS.

ABSTRACT

A large fraction of the baryons in the Universe must be dark. Some of the dark baryons may be in the form of a hot intergalactic medium, while others may have been processed into the remnants of a first generation of Population III stars. Such remnants might be numerous enough to provide the dark matter in galactic halos, although halos could also contain elementary particles relics of the Big Bang. The remnants would need to be either brown dwarfs or massive black holes. Evidence for the first possibility may come from cluster cooling flows and gravitational microlensing events. Evidence for the second could come from detecting the background light generated by the stellar precursor of the black holes. However, the COBE results imply that the light would need to be in the near-IR rather than being reprocessed by dust into the far-IR.

1. INTRODUCTION

It is well known that, while ordinary visible material has a density $\Omega_V \simeq 0.01$ in units of the critical density, there is evidence for a much larger density of invisible material[1]. In fact, there are four contexts in which the existence of dark matter has been proposed: (i) There may be *local* dark matter[2], associated with our galactic disk, with a mass comparable to that of the visible disk, though the need for this may go away if there is a population of "thick disk" stars[3]. (ii) There may be dark matter in galactic *halos*[4]; the associated density parameter is rather uncertain since it depends on the unknown radius to which the typical halo extends (averaged over all galaxies) but it is probably of order $\Omega_h \simeq 0.1$. (iii) There may be dark matter in *clusters*, with a density parameter in the range $\Omega_c \simeq 0.2$-0.3, depending on the scale of clustering. (iv) Finally, if one accepts the inflationary scenario, there may have to be *background* dark matter (unclustered on the scale of galaxies though probably clumped on smaller scales) in order to make the total cosmological density parameter unity.

Some of the dark matter components may be the same. For example, if one believes that individual galaxies are stripped of their halos when they aggregate to form clusters (thereby forming a collective cluster halo), it would be fairly natural to identify the halo and cluster dark matter providing the original halos were large enough. Likewise the cluster and background dark matter could be identified if one invoked some form of biased galaxy formation in which galaxies form in only a small fraction of the volume of the Universe[5]. On the other hand, it is also possible that all the dark matter components are different or that they require different mixtures of dark matter candidates.

Candidates for the dark matter may be grouped into *non-baryonic* types (in which the dark object is some sort of elementary particle) and *baryonic* types (in which it is something astrophysical). In the first case, the existence of the dark object - which may be generically termed an "ino" - goes back to the very early Universe. In the second case, the dark object forms out of the background gas at a relatively late stage (viz. 10^7-10^9y after the Big Bang); this may be termed the "Population III" scenario. The candidates are listed explicitly in Table (1) in order of increasing mass. The table illustrates that there are many forms of non-luminous matter; even though some of the candidates can be rejected,

Table (1): baryonic and non-baryonic dark matter candidates

INOS		POPULATION III	
Axions	$(10^{-5} eV)$	Snowballs	?
Neutrinos	$(10\ eV)$	Brown dwarfs	$(< 0.08\ M_\odot)$
Photinos	$(1\ GeV)$	M-dwarfs	$(0.1\ M_\odot)$
Monopoles	$(10^{16} GeV)$	White dwarfs	$(1\ M_\odot)$
Planck relics	$(10^{19} GeV)$	Neutron stars	$(2\ M_\odot)$
Primordial holes	$(>10^{15} g)$	Stellar holes	$(\sim 10\ M_\odot)$
Quark nuggets	$(<10^{20} g)$	VMO holes	$(10^2 - 10^5 M_\odot)$
Shadow matter	?	SMO holes	$(>10^5 M_\odot)$
Cosmic strings	?		

several viable ones in both categories remain. It is therefore naive to assume that *all* the dark matter problems will have a single explanation and one probably needs *both* baryonic and non-baryonic components.

The main argument for both baryonic and non-baryonic dark matter comes from cosmological nucleosynthesis considerations. In order to explain the primordial light element abundances, the latest calculations[6] – allowing for a recent reduction in estimates of the neutron lifetime – require the baryon density parameter to lie in the range $0.01h^{-2} < \Omega_b < 0.016h^{-2}$ (where h is the Hubble parameter H_0 in units of 100 km/s/Mpc). Since H_0 must be at at least 50, the upper limit implies that Ω_b must be well below 1, so the inflationary model requires that the "smooth" background density be dominated by some non-baryonic component. Whether Ω_b exceeds the density of visible baryons $(\Omega_v \sim 0.01)$ depends crucially on the value of H_0. Most of the baryons could be visible if $H_0 = 100$ but, in this case, the age of the Universe would be less than the ages of globular clusters unless there were a cosmological constant. On the other hand, if $H_0 = 50$ (as seems more likely), Ω_b must exceed Ω_v by a factor of at least 4, so there must be *some* dark baryons and perhaps enough to provide galactic halos.

Much recent work has focussed on the question of whether one can circumvent the cosmological nucleosynthesis upper limit on Ω_b by invoking a 1st order phase transition during the quark-hadron era[7]. The idea is that the transition would generate fluctuations in the baryon density. Neutrons would then diffuse from the overdense regions (since their

cross-section is less than that of the protons), leading to variations in the neutron-proton ratio[9]. It is conceivable that $\Omega_b=1$, in which case 99% of baryons have turned dark[9]. However, this is not a view favoured by most cosmologists and I will not discuss it further here.

The discrepancy between Ω_b and Ω_v could be resolved if there were an appreciable density of intergalactic gas. We know that there must be some neutral gas in the form of Lyman-α clouds but the density parameter associated with these clouds is probably no more than 0.01. If one wants to put all the dark baryons into an intergalactic medium, then its temperature must lie in the range 10^4K to 10^8K, the lower limit coming from the Gunn-Peterson test and the upper limit from the new COBE limits on the Compton distortion of the microwave background radiation[10].

The other possibility is that the dark baryons are in Population III remnants. Note that this does not exclude the first proposal: in a biased scenario, for example, dark baryons could be in *both* galactic halos and an intergalactic medium, in which case inos could still *dominate* the halo density. However, most of the emphasis in this talk will be on the possibility that halos are dominated by Population III objects. The suggestion that the halo dark matter could be baryonic goes against the current trend to assume that all forms of dark matter except that in the disk are non-baryonic. However, the arguments advanced in support of this trend[11] are not very convincing but just reflect a prejudice that the number of forms of dark matter should be as small as possible. Table (1) indicates that dark matter could take as many different forms as visible matter, so it is quite plausible that the efficiency with which baryons turn dark exceeds the efficiency with which they turn visible. Thus the fact that the background dark matter (if such exists) has to be non-baryonic does not exclude the halo dark matter being baryonic. Admittedly, it might seem strange that baryonic and non-baryonic material should have comparable densities, although there are some scenarios where this arises naturally[12], but this is a coincidence which pertains independent of whether the baryons remain in mainly visible or invisible form.

In the next two sections I will discuss why one might expect Population III objects to form (§2) and what constraints observations already place on their nature (§3). The remaining sections will discuss recent developments: in particular, the implications of the COBE results (§4), the evidence for dark matter production in cooling flows (§5) and the detection of dark matter through gravitational lensing effects (§6).

2. THE FORMATION OF POPULATION III OBJECTS

Although the halo and possibly cluster dark matter may be baryonic, it cannot be in the form of ordinary gas else it would generate too many X-rays. The gas must therefore have been processed into the dark remnants of a generation of pregalactic or protogalactic "Population III" stars[13]. The reason one might expect Population III stars to form is that, in most cosmological scenarios, one would expect the first bound objects to be much smaller than galaxies. For example, in the *hierarchical clustering* scenario, the first objects have a mass around $10^6 M_\odot$ and bind at a redshift in the range 20-100; these clouds then cluster gravitationally to make galaxies and clusters of galaxies[14]. A currently popular version of this model is the "Cold Dark Matter" scenario[15]. In the *pancake* scenario[16] the first objects to form are pancakes of cluster or supercluster scale and they do so at a redshift in the range 3-10. They first fragment into clouds of 10^6-$10^9 M_\odot$ and so these clouds have to cluster before galaxies can form. One version of this theory is the "Hot Dark Matter" scenario[17], although this is no longer very popular because galaxies probably form too late.

In all these scenarios, an appreciable fraction of the Universe may go into pregalactic clouds before galaxies themselves form. The question then arises of what happens to these clouds. In some circumstances, one expects them to be disrupted by collisions with other clouds because the cooling time is too long for them to collapse before such collisions. However, there is usually some mass range in which the clouds cool fast enough to survive. For example, the range is 10^6-$10^{11} M_\odot$ in the CDM scenario. In this case the clouds could face various possible fates. They might just turn into ordinary stars and form objects like globular clusters. On the other hand, the conditions of star formation could have been very different at early times and several alternatives have been suggested:

* Some people argue that the first stars could have been much smaller than at present, either because of the enhanced formation of molecular hydrogen at early epochs[18] or because of the prevalence of high pressure pregalactic cooling flows[19]. Such cooling flows would be analagous to the cluster flows observed at the present epoch[20] but on a smaller scale.

* Other people argue that the first stars could have been much larger than at present. For example, the fragment mass could be increased before metals formed because cooling would be less efficient[21]. There is also observational evidence that the IMF may become shallower as metallicity decreases[22], thereby increasing the fraction of high mass stars. Another possibility is that the effects of the microwave background could increase the characteristic fragment mass[23].

* One could also get a mixture of small and large stars. For example, one proposal[24] is that one gets the formation of massive exploding stars in the core of the cloud, followed by the formation of low mass stars where the gas swept up by the explosions encounters infalling gas. Alternatively, one could envisage a scheme in which tidally induced angular momentum effects lead to a disk of small stars around a central very massive star[23].

* It is possible that the first clouds do not fragment at all. For example, they might collapse directly to supermassive black holes or they might remain in purely gaseous form and become Lyman-α clouds[25]. In the latter case, the formation of the dark-matter-producing would need to be postponed until the epoch of galaxy formation.

This discussion indicates that, while there is considerable uncertainty as to the fate of the first bound clouds, they are likely to fragment into stars which are either larger or smaller than the ones forming today. (People merely disagree about the direction!) One certainly *needs* the stars to be different from the ones forming now if they are to produce a lot of dark matter. One also requires the clouds to fragment very efficiently. At first sight, this might seem rather unlikely but there are circumstances even in the present epoch were this occurs: for example, in starburst galaxies or cooling flows.

We note that there is no necessity for the Population III stars to form *before* galaxies just as long as some change in the conditions of star formation alters the mass spectrum. However, the epoch of formation will be very important for the relative distribution of baryonic and non-baryonic dark matter, especially if the non-baryonic dark matter is "cold" so that it can cluster in halos. This would apply, for example, if the ino was a "Weakly Interacting Massive Particle" or WIMP. In this case, if

the Population III stars form *before* galaxies, one might expect their remnants to be distributed throughout the Universe[26] with the ratio of the non-baryonic and baryonic densities being the same everywhere and of order 10. If they form at the same time as galaxies, perhaps in the first phase of protogalactic collapse, one would expect the remnants to be confined exclusively to halos and clusters. In this case, their contribution to the halo density could be larger since the baryons would probably dissipate and become more concentrated. If the inos are hot and cannot cluster in halos, then halos would consist exclusively of machos. These possibilities are illustrated in Figure (1). Greist has coined the term "Massive Compact Halo Object" or MACHO for the baryonic dark object, so the main message of Figure (1) is that we may need both machos and wimps!
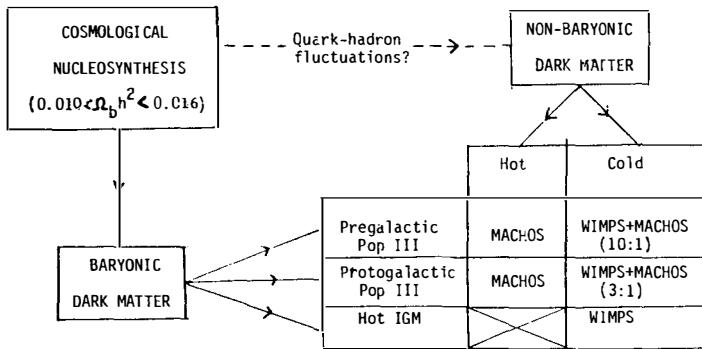
Figure (1): This indicates schematically the relative contribution to galactic halos from baryonic and non-baryonic dark matter in various scenarios. The ratio depends upon whether the non-baryonic dark matter is hot or cold and on whether the Population III objects form at a pregalactic or protogalactic epoch.

3. CONSTRAINTS ON POPULATION III OBJECTS

We have seen that it is difficult to predict *a priori* the mass of the Population III objects. However, Table (1) shows that there is a wide range of masses over which stars are expected to leave dark remnants. The suggestions range from objects as small as snowballs to objects as large as supermassive stars. We will discuss each of the possibilities in turn, although we will find that many of them can be rejected on empirical grounds.

Snowballs. Condensations of cold hydrogen have been proposed but can be probably be excluded[11]. In order to avoid being disrupted by collisions within the age of the Universe, they must have a mass of at least 1 g. In addition, objects in the mass range 10^{-3}g to 10^7g are excluded from providing either the local or halo dark matter by the observed upper limit on the frequency of encounters with interstellar meteors. Those in the mass range 10^7g to 10^{16}g are excluded by the number of impact parameters on the Moon, while those in the mass range 10^{15}g to 10^{22}g are excluded by the fact that no interstellar comet has crossed the Earth's orbit in the last 400 years. These limits are discussed in detail elsewhere[27]. Although there are no empirical reasons for excluding snowballs larger than 10^{22}g, it is difficult to envisage a scenario in which a cloud would fragment into objects smaller than (say) the mass of Jupiter, so snowballs are not a very plausible dark matter candidate. Note that the argument[11] that snowballs would be evaporated by the microwave background radiation is probably incorrect[28].

Brown dwarfs. These are objects smaller than 0.08 M_Θ – sometimes termed "jupiters" – which never ignite their nuclear fuel. Fragmentation could in principle lead to objects as small as this[18]. If the disk dark matter comprises such objects, the nearest one would be at a distance of about $6000(M/M_{Jup})^{1/3}$AU; this distance would increase by a factor of 2 for halo objects. Such objects might conceivably be detectable as infrared sources, although analysis of IRAS data has found no evidence for them. The best way to detect brown dwarfs in the halos of other galaxies may be by their gravitational microlensing effects[29]. This is because, if a galaxy is positioned so as to image-double a distant quasar, then there is also a high probability that an individual halo object will traverse the line of sight of one of the images. This will give appreciable intensity fluctuations in one but not both images. The effect would be observable for objects larger than $10^{-4}M_\Theta$ but the timescale of the fluctuations, being of order $40(M/M)^{1/2}$y, would only show up over a reasonable period for M<0.1 M_Θ. There is also the possibility of looking for the very rare microlensing events associated with objects in our own halo.

M-dwarfs. Nuclear-burning stars would have to be as small as 0.1 M_Θ in order to have a mass-to-light ratio large enough to explain any of the dark matter problems. However, such low mass stars could still be detectable as infrared sources and searches already indicate[30] that their number density near the Sun can be at most 0.01 pc^{-3}. This is a hundred

times too small to explain the local dark matter problem and ten times too small to explain the halo problem. Infrared observations[31] also indicate that the halos of other spiral galaxies cannot consist of M-dwarfs. We therefore reject this possibility.

White dwarfs. These would be the natural end-state of stars with initial mass in the range 1-4 $M_\odot$ and they could certainly fade below detectability within the age of the Galaxy[32]. The fraction of the original star which is left in the white dwarf remnant is rather low but one could still have a large remnant density if there were many generations of stars[33]. The problem is that their precursors would also make a lot of helium. Despite claims to the contrary[34], this almost certainly precludes their density being high enough to explain the dark matter in halos or clusters, although it might be large enough to explain the local dark matter.

Neutron stars. These would be the natural end-state of stars in some mass range above 8 $M_\odot$. However, such precursors would also return a large fraction of their mass to the background medium as heavy elements[35]. The fact that the poorest Population I stars have metallicity of order 10^{-3} therefore places an upper limit on the fraction of the Universe's mass which can have been processed through such precursors[13]. This limit precludes neutron stars from explaining any of the dark matter problems.

Stellar black holes. Stars larger than some critical mass M_{BH} may leave black hole rather than neutron star remnants. The value of M_{BH} is uncertain[36] but it is probably around 50 $M_\odot$. In this case, some of the heavy elements generated by nucleosynthesis may be swallowed rather than ejected but one would still expect some mass ejection, so one could not process a large fraction of the mass of the Universe into the holes. Note that stellar black holes can probably not provide the disk dark matter. This is because the survival of binaries in the disk may require[37] that the local dark objects be smaller than 2 $M_\odot$, although this limit has been disputed[38].

VMO black holes. Stars in the mass range above 100 $M_\odot$, which are termed "Very Massive Objects" or VMOs, would experience the pair-instability during their oxygen-burning phase[39]. This would lead to disruption below some critical mass M_C but complete collapse above it[40]. VMO black holes are therefore more plausible dark matter candidates than ordinary stellar black holes, although they do not seem to form at the

present epoch. In the absence of rotation $M_c \simeq 200$ $M_\odot$ but it could be as high as $2 \times 10^4 M_\odot$ if rotation were maximal[41]. An important constraint on the number of VMO remnants is provided by background light limits. However, this constraint is sensitive to the redshift of star formation, as discussed later. VMO black holes could also be detected by gravitational lensing effects. This is because such objects would modify the ratio of the line to continuuum output of a quasar[42]; the fluxes are affected differently because they come from regions which act as extended and pointlike sources, respectively. This already excludes black holes in the mass range 10^2-$10^5 M_\odot$ from having a critical density, though not necessarily the tenth critical density required for galactic halos. Note that laser interferometry could eventually detect the gravitational wave background generated by VMO black holes, especially if they form in binary systems[43].

SMO black holes. Stars larger than $10^5 M_\odot$ are termed "Supermassive Objects" or SMOs and would collapse directly to black holes without any nuclear burning if they have zero metallicity due to relativistic instabilities[44]. Such holes are already supposed to reside in galactic nuclei, although those ones would only have a tiny cosmological density. As discussed later, halo black holes would heat up the disk stars in our own galaxy more than is observed unless they were smaller than about $10^6 M_\odot$, so they would have to lie in the narrow mass range 10^5-$10^6 M_\odot$. Even if the dark matter in clusters is different from that in halos, the absence of unexplained tidal distortion in the visible galaxies implies[45] that the cluster objects must be smaller than $10^9 M_\odot$. The number of SMO holes is also constrained by gravitational lensing effects[46]. If one has a population of objects with mass M and density parameter Ω, then the probability of one of them image-doubling a distant quasar is Ω and the separation between the images is $10^{-6}(M/M_\odot)^{1/2}$ arcsec. Thus VLBI, with a resolution of 10^{-3} arcsec, could search for objects as small as $10^6 M_\odot$, while limits from the VLA, with a resolution of 0.1 arcsec, already imply that holes bigger than $10^{11} M_\odot$ must have less than 0.4 of the critical density[47]. Note that the background gravitational waves produced by the formation of SMO holes could in principle be detected by the Doppler tracking of interplanetary spacecraft[48].

The various constraints on the form of the dark matter discussed above are brought together in Table (2). This indicates the viable explanations for each of the dark matter problems. The shaded regions are excluded by either dynamical, nucleosynthetic, lensing or point-source

Table (2): This shows the constraints on the possible dark matter candidates, shaded regions being excluded by at least one of the limits discussed in the text. SMO, VMO and BH refer to the black hole remnants of Supermassive Objects, Very Massive Objects and ordinary stars; NS=neutron star; WD=white dwarf; MD=M-dwarf; BD=brown dwarf; PBH=primordial black hole (here regarded as a non-baryonic candidate).
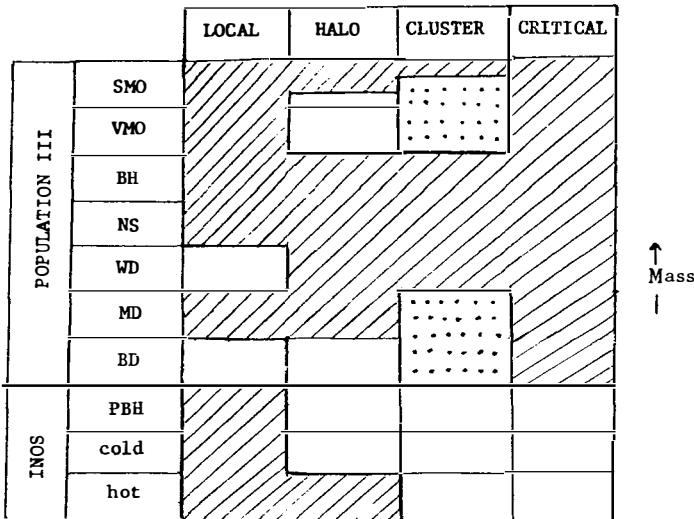
limits. Whether the dotted region is excluded depends on whether the cosmological nucleosynthesis constraint requires that the cluster dark matter be non-baryonic, which is marginal even in the standard scenario. Although a critical density of baryons might be permitted in non-standard scenarios, this is excluded in most mass ranges anyway. The constraints on the ino candidates are all dynamical in nature and depend upon whether the particles are hot or cold. None of the inos could be expected to provide the local dark matter since they are non-dissipative and it seems difficult for hot particles to cluster enough to explain galactic halos[49].

The prime message of Table (2) is that one could not expect any single candidate to explain all four dark matter problems. On the other hand, the table does constrain the possible solutions: (i) the local dark matter - if it exists - would have to be white dwarfs or brown dwarfs; (ii) the halo dark matter could be either brown dwarfs or black holes or inos; (iii) the cluster dark matter would need to be inos if one adopts the cosmological

nucleosynthesis limit in its strongest form but the other halo candidates would be viable otherwise; (iv) the background dark matter – if it exists – would have to be inos unless one gives up the cosmological nucleosynthesis limit altogether. Note that the halo and cluster dark matter could well be a *mixture* of the different candidates, as indicated in Table (1). Having a mixture need not imply any obvious dynamical signature, especially if the inos (like the Population III remnants) are cold.

4. VMOS AND BACKGROUND LIGHT

The most direct evidence for a population of VMOs would come from the detection of the background light they generate. Each VMO has a luminosity $L \simeq 10^{40}(M/10^2 M_\odot)$ erg/s and a surface temperature of about 10^5K (independent of the mass M). In the absence of dust absorption, one would therefore expect a background with a present density and peak wavelength given by

$$\Omega_R = 4 \times 10^{-6} \left[\frac{\Omega_*}{0.1}\right] \left[\frac{1+z_*}{100}\right]^{-1} , \quad \lambda = 4 \left[\frac{1+z_*}{100}\right] \mu. \tag{1}$$

Here Ω_R is measured in units of the critical density, Ω_* is the density of the stars in the same units, and z_* the redshift at which they burn. Comparison with the upper limits on the background radiation density in the near-IR to UV bands[50] shows that VMOs with the density $\Omega_* \simeq 0.1$ required to explain galactic halos would have to burn at a redshift exceeding 30; otherwise Ω_R would be too large. This would imply that they were necessarily pregalactic. However, if *dust* were present, the radiation could be reprocessed to longer wavelengths, where the limits on the background density are weaker. In this case, the re-emitted radiation should have spectrum peaking at a present wavelength[51]

$$\lambda_{peak} = 700 \left[\frac{\Omega_R h^2}{10^{-6}}\right]^{-1/5} \left[\frac{r_d}{0.1\mu}\right]^{1/5} \left[\frac{1+z_d}{10}\right]^{1/5} \mu \tag{2}$$

where r_d is the grain size, z_d is the redshift of light production or grain production (whichever is smaller) and we have assumed that the grain opacity scales as λ^{-1}. For comparison, the microwave background peaks at 1600μ. The total background spectrum should thus have three parts: the microwave background component, the far-IR or submillimetre dust component, and the attenuated starlight component.

There have, in fact, been two claimed detections of a far-IR background. The first[52] was based on IRAS data and suggested a background with $\Omega_R \simeq 3 \times 10^{-6} h^{-2}$ at 100μ. However, this was always controversial, since it required subtracting the (very uncertain) zodiacal emission, and the current upper limit is a factor of 3 below this. More recently, a significant excess in the spectrum of the microwave background radiation was reported in the submillimetre waveband, following a rocket experiment by a team from Nagoya and Berkeley[53]. The background appeared to peak at around 700μ with a density of $\Omega_R \simeq 5 \times 10^{-6} h^{-2}$. This was very exciting since it seemed to confirm the VMO prediction.

Unfortunately, the Nagoya-Berkeley excess has now been disproved by COBE[54]. However, this need not exclude the VMO scenario itself since one can always argue that there is not enough dust to reprocess the radiation. For example, *pregalactic* dust can only absorb the VMO light for

$$1 + z_d > 8 \left[\frac{\Omega_d h}{10^{-5}} \right]^{-2/3} \left[\frac{r_d}{0.1\mu} \right]^{2/3} \tag{3}$$

where Ω_d is the dust density parameter, normalized to the value associated with galaxies since this is presumably an upper limit to the pregalactic abundance. It is not inevitable that condition (3) should be satisfied; indeed there need not be any pregalactic dust at all. In order for *galactic* dust to absorb the VMO light, the redshift of galaxy formation (z_G) must satisfy two independent conditions: firstly, it must exceed the value given by eqn (3) with Ω_d normalized to the mean galactic value; secondly, one needs the galaxies to cover the sky, which – for galaxies like our own – requires

$$1 + z_G > 10 \left[\frac{R_G}{10 \text{kpc}} \right]^{2/3} \left[\frac{\Omega_G}{0.01} \right]^{-2/3} \tag{4}$$

Here R_G is the radius and Ω_G is the baryonic density associated with the galaxies. It is by no means clear that galaxies formed this early and they would certainly not do so in the CDM scenario. We conclude that a halo density of VMOs could still exist but, if so, their light must be in the near-IR[56]. In fact, the Japanese have also claimed to detect a background at 2μ with density $\Omega_r = 2 \times 10^{-6} h^{-2}$ and this is certainly compatible with equation (1). They claim the background is very narrow and suggest that it is a redshifted Lyman-α line. In this case it would need to originate at $z_* = 16$. If 2μ is just the original black-body peak, eqn (1) implies $z_* = 50$.

It should be noted that some people have proposed that the *entire* microwave background is grain-thermalized starlight[57]. This is possible in principle but the grains would have to form at a very high redshift (z>200) and be very elongated in order to thermalize at long wavelengths[58]. In this case, one would also need to invoke VMOs to generate the observed primordial helium abundance, since the standard cosmological nucleosynthesis picture no longer applies. Although VMOs might naturally generate a helium abundance of about 25%, one still needs to generate the other light element abundances[59], so this scenario is not as attractive as the standard one.

5. BROWN DWARFS AND COOLING FLOWS

X-ray observations show that many clusters contain hot gas with a temperature of about 10^8K. In clusters dominated by a central cD galaxy, the emission is peaked at the core, indicating a high central gas density. Since the associated cooling time is less than the Hubble time, one expects the gas in the core to be flowing inwards, driven by the pressure of the surrounding gas, which is too tenuous to cool appreciably. The observational evidence for such cooling flows has been reviewed elsewhere[20]. The mass flow rate associated with the flows varies from a few $M_\odot$ y^{-1} to $10^3 M_\odot$ y^{-1} and the mass appears to be deposited over a wide range of radii. However, the gas cannot be forming stars with the same mass spectrum as the solar neighbourhood, else the central region would be bluer and brighter than observed. This suggests that the cooling flows produce very low mass stars, possibly because the high pressure reduces the Jeans mass (assumed to provide an upper limit to the mass of the stars being formed). The important feature of a cooling flow is that it is *quasi-static*, in the sense that the cooling time exceeds the dynamical time, and it is this condition which is supposed to preserve the high pressure.

Although cooling flows provide a natural way of turning gas into low mass stars with high efficiency, those observed at the present epoch are mainly confined to the central galaxies in clusters and therefore could not in themselves be responsible for either the cluster dark matter (since this is distributed throughout the cluster) or the halo dark matter in galaxies outside clusters. However, one could expect analagous high pressure quasi-static flows to occur at earlier cosmological epochs and these would

have been on much smaller scales than clusters[19]. This conclusion pertains in either the hierarchical clustering or pancake scenario, although we will only focus on the hierarchical case here.

The crucial point is that the amount of mass cooling quasi-statically in a gravitationally bound cloud is maximimized when the cooling time t_c is comparable to the free-fall time t_f: collapse does not proceed at all for $t_c > t_f$, whereas it is not quasi-static for $t_c < t_f$. Now in any particular variant of the hierarchical clustering scenario, one can specify the mass binding as a function of redshift. If we focus on a cloud of mass M, the dynamical time will just be of order the Hubble time at that redshift, whereas the cooling time will depend upon the density and virial temperature of the cloud (which are themselves determined by M and z). Thus one can specify a region in the (M,z) plane in which bound clouds will cool within a dynamical time. This applies above a *lower* mass limit associated with molecular hydrogen or Lyman-α cooling and below an *upper* mass limit is associated with atomic hydrogen cooling. The condition $t_c \sim t_f$ will be satisfied at the boundary of the region and the intersection of this boundary with the binding curve M(z) singles out two characteristic mass-scales and redshifts. These correspond to what we term "pervasive pregalactic cooling flows" (PPCFs) since the amount of gas processed through the cooling flows is maximimized. The associated mass-scales are always of order $10^4 - 10^8 M_\odot$ and $10^{11} M_\odot$ but the redshifts depend on the particular scenario.

The question of the scale of the cooling flows is very important because this should also be the scale on which the dark matter aggregates today. If the scale is of order $10^{11} M_\odot$, then cooling flows could make galactic halos directly. If it is of order $10^6 M_\odot$, then the first objects to form would be dark clusters of this mass. Galactic halos would then form as a result of the agglomeration of these objects. One might anticipate *most* of the dark matter being made on the smaller scale since most of the gas will have been consumed by the time atomic cooling becomes important. However, this may not apply in the Cold Dark Matter picture because the spectrum of fluctuations is very flat on subgalactic scales. Furthermore, in some versions of the hierarchical clustering picture, M(z) is never small enough for low mass PPCFs to occur[60]. This suggests that most of the dark matter would need to be made by high mass PPCFs. However, the problem with this is that most of the baryons will by then have gone into clouds with $t_c < t_f$ and such clouds should *not* make low mass stars.

One way round this is to argue that even clouds with $t_c < t_f$ initially can make a lot of dark matter. The idea is that gas always drops out at such a rate as to preserve the PPCF condition $t_c \sim t_f$. One thus gets a two-phase medium, with cool dense clouds embedded in hot high pressure gas. This was originally proposed as a mechanism to make globular clusters at a protogalactic epoch[61] but one can show that sufficiently small clouds would tend to fragment into dark clusters rather than visible clusters, at least in the presence of molecular hydrogen[62].

The possibility that galactic halos could comprise dark clusters would have interesting dynamical implications. Earlier we mentioned that the halo objects must be smaller than $10^6 M_\odot$ else they would overheat the stars in the stellar disk. More positively, it has been have proposed[63] that halo objects of mass $2 \times 10^6 M_\odot$ could be *required* in order to generate the amount of observed disk-heating. In particular, this would explain why the stellar velocity dispersion increases with stellar age as $t^{1/2}$ and why the radial, azimuthal and vertical dispersions have their observed ratios. It was originally proposed that the halo objects are single supermassive black holes but this has problems: too many holes tend to drift into the galactic nucleus as a result of dynamical friction, while the ones which stay in the halo tend to generate too much radiation through accretion as they traverse the disk[64.]

One way to circumvent these problems is to assume that the disk heaters are $2 \times 10^6 M_\odot$ clusters of smaller objects[65]. The accretion luminosity is then reduced by a factor which is of order the number of objects per cluster and the dynamical friction problem is avoided provided the clusters are disrupted by collisions before they reach the galactocentric radius of 2 kpc at which they become affected by the dynamical friction of the Galactic Spheroid. This requires that they have a radius r_c exceeding about 0.1 pc. On the other hand, in order to explain the observed disk-heating down to 4 kpc, we need $r_c < 4$ pc. Thus r_c is rather tightly constrained. The upper limit also requires the cluster components to be smaller than 10 $M_\odot$ (else they would evaporate too quickly through 2-body relaxation) and this suggests that they must be brown dwarfs rather than black holes.

6. GRAVITATIONAL LENSING

One of the most useful signatures of Population III objects would be their gravitational lensing effects. Indeed it is remarkable that lensing could permit the detection of Population III objects over the entire mass range $10^{-9}M_\odot$ to $10^9 M_\odot$. There are three distinct types of lensing effect and we saw in §3 how each of these can be used to constrain the possible baryonic candidates.

The multiple-imaging of a distant source like a quasar is termed "macrolensing" and constraints on the frequency of such multiple-images can be used to limit the density of SMOs. In fact, the data on macrolensing by galaxies already provide evidence for dark matter in galactic halos since the image separation usually requires a lensing mass exceeding the mass associated with the visible galaxy. In the case of 2016+112, for example, the lens model requires that the overall mass-to-light ratio be at least 100 and that the total matter distribution be different from the visible distribution[66]. Macrolensing may also provide evidence for dark matter in clusters. For example, the giant arc A370 is thought to be a galaxy lensed by a rich cluster and, in this case, the cluster must have a dark mass of $3 \times 10^{14} M_\odot$, well in excess of the visible mass[67].

We saw how the absence of variations in the line-to-continuum ratio for different images of a lensed quasar can be used to constrain the density of VMOs. Positive evidence for such an effect could actually indicate the nature of the dark matter. Such evidence may already exist in the case of the double quasar 2016+112, where the different intensity ratios for the images suggest[68] that the lensing objects have a mass in the range $3 \times 10^4 M_\odot$ to $3 \times 10^7 M_\odot$. It is interesting that clusters in this mass range are a natural consequence of the pregalactic cooling flow scenario, although they could also arise in other ways.

Intensity variations in one image of a lensed quasar due to individual halo objects within the lensing galaxy is termed "microlensing" and we saw how this effect may be used to constrain the density of low mass stars. Again, there may be positive evidence for this effect since lensing by jupiters may explain the existence of some optically violently variable quasars[69]. Rather more convincing evidence comes from a direct detection of microlensing[70] for the quasar 2237+0305. This has four images at a redshift of 1.7 and the lens is a galaxy at redshift 0.04. The brightest image brightened by 0.5^m from September 1987 to August 1988 and then

dimmed by 0.15^m by September 1988. There was no variation in the other images, even though the difference in light travel-time was only hours. The observed timescale for the variation would require a lensing mass in the range $0.001\ M_\odot$ to $0.1\ M_\odot$. One problem with interpreting this as evidence for dark matter is that the variable image is almost exactly aligned with the centre of the lensing galaxy and here the density should be dominated by ordinary stars rather than dark matter.

Attempts to detect microlensing by objects in our own galactic halo by looking for intensity variations in stars in the Large Magellanic Cloud[71] are already underway and are reported upon elsewhere in this volume. In this case, the timescale for the variation is smaller, so one can seek for lensing masses all the way down to $10^{-9} M_\odot$. However, the probability of an individual star being lensed is very low, so one has to look at many stars for a long time.

7. CONCLUSIONS

We have seen that there are good reasons for believing that many of the baryons in the Universe have gone into dark remnants, although it is unclear whether brown dwarfs or black holes are more plausible. The best signature for black holes would be the background light generated by their stellar precursors but the COBE results remove what might have been good evidence for this. The claim that cooling flows make low mass stars would seem to favour the brown dwarf option. However, given the uncertainty, it is probably best to hedge one's bets until future observations of microlensing or the near-IR background settle the issue. Of course, it is possible that both options are wrong since we still cannot exclude the dark baryons being in the form of hot intergalactic gas. Finally, it should be re-emphasized that the background dark matter, required to make up the critical density if one accepts inflation, must be an ino if one accepts the standard cosmological nucleosynthesis picture. Dynamical considerations suggest that it is probably a cold particle like the photino, in which case the halo and cluster dark matter could well be a *mixture* of baryonic and non-baryonic dark matter.

REFERENCES

1. Faber, S.M. & Gallagher, J.S. (1979). *Ann.Rev.Astron.Astrophys.*, 17, 135.
2. Bahcall, J.N. (1984). *Astrophys.J.*, 276, 169.
3. Gilmore, G. & Kuijken, J. (1989). *Mon.Not.R.Astron.Soc.*, (in press).
4. Trimble, V. (1987). *Ann.Rev.Astron.Astrophys.*, 25, 425.
5. Kaiser, N. (1984). *Astrophys.J.Lett.*, 284, L9.
6. Olive,K.A., Schramm,D.N., Steigmann, G. & Walker,T. (1990). *Phys.Lett.B.*, 236, 454.
7. Applegate, J. & Hogan, C.J. (1985). *Phys.Rev.D.*, 31, 3037.
8. Alcock,C., Fuller, G.M. & Mathews, G.J. (1987). *Astrophys.J.*, 320, 439.
9. Malaney, R.A. & Fowler, W.A. (1988). *Astrophys.J.*, 333, 14.
10. Bartlett, J. (1990). Preprint.
11. Hegyi, D.J. & Olive, K.A. (1986). *Astrophys.J.*, 303, 56.
12. Turner, M.S. & Carr, B.J. (1987). *Mod.Phys.Let.A.*, 2, 1.
13. Carr, B.J., Bond, J.R. & Arnett, W.A. (1984). *Astrophys.J.*, 277, 445.
14. Peebles, P.J.E. & Dicke, R.H. (1968). *Astrophys.J.*, 154, 891.
15. Blumenthal, G.R., Faber, S.M., Primack, J.R. & Rees, M.J. (1984). *Nature*, 311, 517.
16. Zeldovich, Ya.B. (1970). *Astr.Astrophys.*, 5, 84.
17. Bond, J.R., Centrella, J., Szalay, A.S. & Wilson, J.R. (1984). *Mon.Not.R.Astron.Soc.*, 210, 515.
18. Palla,F., Salpeter,E.E. & Stahler,S.W. (1983). *Astrophys.J.*, 271, 632.
19. Ashman, K.M. & Carr, B.J. (1988). *Mon.Not.R.astr.Soc.*, 234, 219.
20. Fabian, A.C., Nulsen, P.E.J. & Canizares, C.R. (1984). *Nature*, 320, 733.
21. Silk,J. (1977). *Astrophys.J.*, 211, 638.
22. Terlevich,R.J. (1985). In *Star Forming Dwarf Galaxies*, p 395, eds. D.Kunth,T.X.Thuan and J.Tran Thanh Van, Editions Frontieres (Paris).
23. Kashlinsky,A. & Rees,M.J. (1983). *Mon.Not.R.astr.Soc.*, 205, 955.
24. Cayrel,R. (1987). *Astr.Astrophys.*, 168, 81.
25. Rees,M.J. (1986). *Mon.Not.R.astr.Soc.*, 218, 25P.
26. White, S.D.M. & Rees, M.J. (1978). *Mon.Not.R.Astron.Soc.*, 183, 341.
27. Hills, J.G. (1986). *Astron.J.*, 92, 595.
28. Phinney, E.S. (1985). *Astrophys.Counter.*, 17, L1.
29. Gott, J.R. (1981). *Astrophys.J.*, 243, 140.
30. Gilmore, G. & Hewett, P. (1983). *Nature*, 306, 669.
31. Boughn,S.P.,Saulson,P.R. & Seldner,M. (1981). *Astrophys.J.Lett.*, 280, L15.
32. Truran, J.W. (1984). *Ann.Rev.Nuc.Part.Phys.*, 34, 53.
33. Larson, R. (1986). *Mon.Not.R.Astron.Soc.*, 218, 400.
34. Ryu, D., Olive,K.A. & Silk, J. (1990). Preprint.
35. Arnett, W.D. (1978).*Astrophys.J.*, 219, 1008.
36. Schild, H. & Maeder, A. (1985). *Astron.Astrophys.Lett.*, 143, L7.
37. Bahcall, J.N., Hut, P. & Tremaine, S. (1985). *Astrophys.J.*, 290, 15.
38. Weinberg, M.D. (1988). *Astrophys.Space Sci.*, 142, 277.
39. Fowler,W.A. & Hoyle,F. (1964). *Astrophys.J.Supp.*, 9, 201.
40. Bond,J.R.,Arnett, W.D. & Carr,B.J. (1984). *Astrophys.J.*, 280, 825.
41. Glatzel,W., El Eid,M.F. & Fricke,K.J. (1985). *Astr.Astrophys.*, 149, 419.
42. Canizares, C.R. (1982). *Astrophys.J.*, 263, 508.
43. Bond, J.R. & Carr, B.J. (1984). *Mon.Not.R.Astron.Soc.*, 207, 585.
44. Fowler,W.A. (1966). *Astrophys.J.*, 144, 180.
45. Van den Bergh, S. (1969). *Nature*, 224, 891.
46. Press, W.H. & Gunn, J.E. (1973). *Astrophys.J.*, 185, 397.
47. Hewitt, J.N. *et al.* (1988). In *Gravitational Lenses*, ed. J.M.Maran, J.N.Hewitt & K.Y.Loh (Springer-Verlag).
48. Bertotti, B. & Carr, B.J. (1980). *Astrophys.J.*, 286, 1000.
49. Tremaine, S. & Gunn, J.E. (1979). *Phys.Rev.Lett.*, 42, 403.

160

50. McDowell, J. (1986). *Mon.Not.R.Astr.Soc.*, 223, 763.
51. Bond,J.R., Carr,B.J. & Hogan,C.J. (1986). *Astrophys.J.*, 306, 428.
52. Rowan-Robinson,M. (1986). *Mon.Not.R.Astr.Soc.*, 219, 737.
53. Matsumoto,T., Hayakawa,S., Matsuo,H., Murakami,H., Sato, S., Lange,A.E. & Richards,P.L. (1988). *Astrophys.J.*, 329, 567.
54. Mather, J.C. *et al.* (1990). COBE preprint.
55. Matsumoto,T.,Akiba,M. & Murakami,H. (1988). *Astrophys.J.*, 332, 575.
56. Bond,J.R., Carr,B.J. & Hogan,C.J. (1990). *Astrophys.J.* (in press).
57. Rees,M.J. (1978). *Nature,* 275, 35.
58. Hoyle, F. (1989). In *Highlights in Gravitation and Cosmology*, p 236, ed. B.R.Iyer, A.Kembhavi, J.V.Narlikar & C.Vishveshwara (Cambridge University Press).
59. Bond,J.R.,Carr,B.J. & Arnett,W.D. (1983). *Nature, 304,* 514.
60. Ashman, K.M. & Carr, B.J. (1990). *Mon.Not.R.Astron.Soc.*, in press.
61. Fall, S.M. & Rees, M.J. (1985). *Astrophys.J.*, 298, 18.
62. Ashman, K.M. (1989). *Mon.Not.R.Astron.Soc.*, in press.
63. Lacey, C.G. & Ostriker, J.P. (1985). *Astrophys.J.*, 299, 633.
64. Ipser, J.R. & Price, R.H. (1975). Astrophys.J.,216, 578.
65. Carr, B.J. & Lacey, C.G. (1987). *Astrophys.J.*, 316, 23.
66. Schneider, D.P. *et al.* (1988). *Astron.J.*, 96, 1755.
67. Soucail, G., Mellier, Y., Fort,B., Hammer, F. & Mathez, G. (1987). *Astron.Astrophys.*, 184, L7.
68. Subramanian,K. & Chitre, S.M. (1987). *Astrophys.J.*, 313, 13.
69. Nottale, L. (1986).*Astron.Astrophys.*, 157, 383.
70. Irwin,M.J., Webster,R.L. & Hewett, P.C. (1989). *Astron.J.*, 98, 1989.
71. Paczynski, B. (1986). *Astrophys.J.*, 304, 1.

SEARCH FOR MACROSCOPIC DARK MATTER
IN THE HALO OF THE MILKY WAY
THROUGH MICROLENSING.
A FEASIBILITY STUDY

Marc Moniez

**Laboratoire de l'Accélérateur Linéaire
Orsay**

ABSTRACT

The possibility of searching for non-visible massive compact objects in the galactic halo is discussed here. The discovery of such objects would solve the problem of the missing mass in the galaxies, and the experiments which investigate for weakly interacting particles assuming a diffuse cloud of dark matter would have to revise their limits. The non-discovery of these objects would exclude the last possibility left for baryonic dark matter, providing good evidence that the galactic halo has to be made of new particles. The description of the general-relativistic microlensing effect and its application to the search of massive compact objects are given here. A feasibility study shows that it is possible to monitor the luminosity of several million stars in the Large Magellanic Cloud with the required precision, in order to detect a possible microlensing phenomenon induced by heavy compact objects (10^{-4} - 10^{-1} solar mass units). A CCD-based experimental setup is described, which would make it possible to search for compact objects in the 10^{-6} - 10^{-4} solar mass unit domain.

1. DARK MATTER AND GALACTIC HALOS

It is widely accepted that the flat rotation curve of stars and interstellar matter around the centre of spiral galaxies provides strong evidence for dark matter halos of yet unknown origin and composition, in which the luminous part of spiral galaxies (such as our own Milky Way) is embedded [1]. We know that these halos are not made of stars or diffuse baryonic matter because they would have been easily detected. The only two remaining possibilities are

- Massive Compact Objects (MCOs) which could be brown dwarves (stars not massive enough to sustain nuclear reactions in their cores), giant-planet-sized objects (Jupiters), smaller planet-like objects, primordial black holes or massive objects made of new particles...

- Weakly Interacting Massive Particles (WIMPs) which could be either neutrinos or photinos (the lightest supersymmetric particles)...

Although the second candidates are generally favoured from galaxy-formation considerations, the first ones can certainly not be ruled out. In particular, the existence of MCOs is favoured by primordial nucleosynthesis results. The MCOs represent the last possibility left for baryonic dark matter.

An experiment sensitive to the gravitational effects of MCOs could decide between these two possibilities. The idea is to survey for several months, at 10% accuracy or better, the luminosity of a large number of stars in the Large Magellanic Cloud (LMC) and search for stars that undergo a characteristic brightening due to a massive compact object passing near the line of sight and producing a gravitational 'microlensing' effect, which is described in more detail below. Whatever the results of this survey, the consequences will be important both for particle physics and astrophysics. If some massive compact objects are indeed found, then this will give a strong indication that there is less room for a diffuse gas of non-baryonic dark matter (like WIMPs) around galaxies. This would mean that present searches for WIMPs are likely to be ineffective, because the WIMPs density would be much lower than assumed in these searches. On the other hand, if such compact objects are not found, then there is a strong incentive for all kind of WIMPs searches.

2. GRAVITATIONAL MICROLENSING

The idea of looking for massive compact objects through this phenomenon was first put forward in a paper by B. Paczynski [2] and was recently reconsidered by C. Alcock during a seminar in Berkeley (Nov. 89).

2.1 Description of microlensing events

When a spatially-small massive object (hereafter called a deflector D) happens to lie close enough to the line of sight between a light source (a star S) and an observer O, then the observer collects more light from the star than he would in the absence of this deflector. Moreover, the light comes to the observer through two separate paths (on either side of the deflector, see figure 1). The closer the deflector is to the line of sight, the larger the amplification is. This phenomenon, known as gravitational microlensing, is due to the general-relativistic deflection of light by massive bodies.

Figure 1 : the deflection of light by a massive object

d_{OD}

d_{DS}

Plane perpendicular
to the line OD
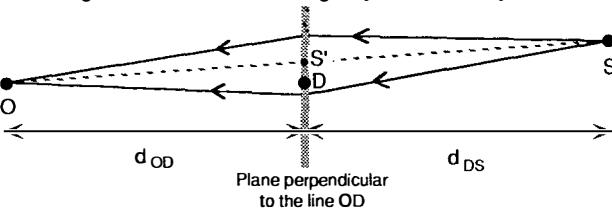
In the special case of perfect alignment of O, D and S, the observer should see a ring of light centred on D (known as an Einstein ring). The radius R_0 of this ring is given by $R_0^2 = 4GMd/c^2$, where G is the Newton constant, M the mass of the deflector, c the speed of light and d a distance such that $1/d = 1/d_{OD} + 1/d_{DS}$. This radius R_0 sets the scale for all distances perpendicular to the line of sight, and is thus a convenient unit. As an example, for a 1 solar mass deflector, situated 10 kpc from the Sun (1 kpc = 1 kiloparsec ~ 3.1 10^{16} km), and for OS = 50 kpc (distance of the Large Magellanic Cloud), the Einstein ring radius is R_0 ~ 1.4 10^9 km.

If the projected position S' of S in the plane perpendicular to OD containing D lies within a circle of radius R_0 centreed on D, then the light from the star S collected by O is amplified by a factor larger than 1.34 (this corresponds to a brightening of the star by more than 0.32 stellar magnitude). Figure 2 shows the amplification factor as a function of the distance DS' (measured in units of R_0). This amplification factor is given by

$$A(u) = \frac{u^2 + 2}{u(u^2 + 4)^{1/2}}, \text{ where } u = DS' / R_0$$
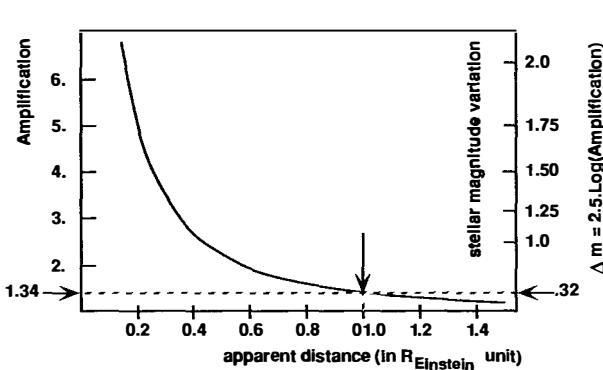
Figure 2 : light amplification versus the distance DS', measured in units of R_0

In any realistic case, the two light rays come to the observer at a relative angle that is very small (less than 0.002 arc second). Thus, the two rays cannot be resolved and the only noticeable effect for the observer is the brightening of the source star.

Given the fact that the alignment has to be very good between the observer, the source and the deflector to get a significant effect, the relative velocity (~ 200 km/s) of these bodies makes the brightening of the star a relatively short time scale phenomenon. For the time and distance scales involved, the relative motion of D with respect to the line OS is linear to a very good approximation. Thus, the time-dependent brightening curve for a given impact parameter u_{min} is easily obtained from the formula for A(u) (see figure 3 taken from reference [2]). The important characteristics of this light curve are its symmetry, uniqueness (the brightening happens only once) and its achromaticity (the brightening is identical, whatever the radiation length emitted by the source), in contrast to all the known light curves for intrinsic variable stars. For brightenings larger than 0.32 magnitude, 56%, 92.5% and 98.5% of the events undergo a magnitude variation smaller than 0.50, 1.00, 1.50 magnitude respectively.

Figure 3 : light curves of a microlensing event
(from reference [2])
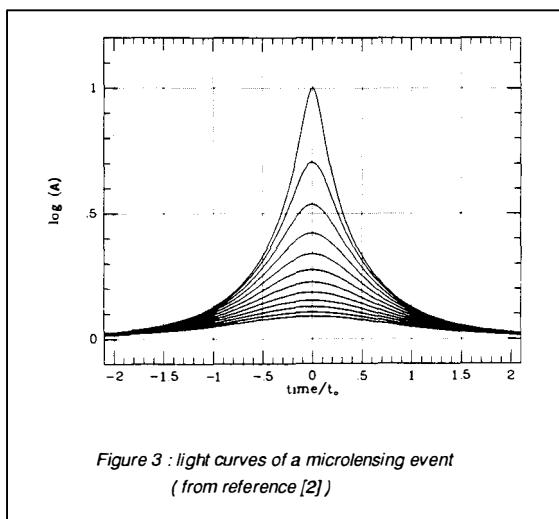
All these characteristics are exact only in the limit that the deflector D and star S are pointlike and simple (no double star, for example). We now examine how they are modified when one of these hypotheses fails.

- D cannot be considered pointlike when its diameter is larger than its Einstein ring radius R_0. In that case, there is an eclipse of the star rather than a brightening. With the same numerical values used above – OD = 10 kpc and OS = 50 kpc – and for a density of 1 g/cm^3 for D, this is the case for deflector diameters smaller than 0.5 km which are objects not accessible in the proposed search.

- S cannot be considered pointlike when its radius, projected in the deflector plane, is comparable to, or larger than $2xR_0$. In this case, there is no impact parameter for which the average amplification is larger than 1.32. In practice, this excludes from the search using LMC the very small deflectors (smaller than 10^{-6} solar mass) that may only be able to significantly amplify a too restricted population of the smallest stars.

- If S is in fact a multiple star system – this is the case for half of the stellar systems – not resolved by the observer, then if the components are very close (projected distance in the deflectors plane d $<<R_0$), the light curve is not affected. If they are very distant (d $>>R_0$), one component will undergo the brightening and the others will remain unaffected. Thus the relative light brightening will be smaller than in the simple case and it may go unnoticed. If d $\sim R_0$, then each component undergo different amplifications, giving a combined brightening curve that is more complicated than the ideal one. Moreover, if these stars have different colours (different surface temperatures), the combined brightening is no longer achromatic.

- If D is in a multiple system, then if the components are very close (d $<<$ individual Einstein radii) then the light curve is not changed, if they are at distances comparable to the individual Einstein radii then the light curve becomes more complicated (not symmetric) but remains achromatic, and if they are very far away then the light curve is the same as when the deflector is simple.

We have also studied the possibility that the deflectors are grouped in clusters of the typical size of a globular cluster (10^5 to 10^6 solar masses). Even in this case, there are enough clusters in the Galactic halo in front of the LMC (at least a few per square degree) and we find that the components of the clusters can be considered as independent deflectors, because their average mutual apparent distance is much larger than their Einstein radii. The clusters as a whole may deflect light from the monitored stars by an amount that is larger than in an individual microlensing event, but this has no observable consequence because this global deflection varies on a time scale much larger than the observing period.

2.2 The expected frequency of microlensing events

The probability that a star undergoes a brightening due to microlensing with an amplification larger than 1.34 is simply the probability that the line OS crosses one of the zones of "gravitational influence" of area πR_0^2 associated with a deflector. Because this area is proportional to the mass of

the deflector, it is easy to see that, at any given moment, this probability is directly proportional to the total mass along the line of sight OS (and independent of the mass distribution of the deflectors). Using a common halo model that explains the observed Galactic rotation curve [3][1], we estimate this probability to be 0.1 10^{-6}, 0.5 10^{-6} and 0.5 10^{-6} respectively for stars at the Galactic centre, in the LMC and in the SMC (Small Magellanic Cloud). For the case of deflectors of 1 M $_\odot$ (one solar mass), figure 4 shows the distribution of the 'transit time', period during which the amplification is larger than 1.34. For deflectors of mass 10^{-2n} M $_\odot$, the horizontal scale of this figure should simply

---

166

be scaled by a factor 10^{-n}. Assuming a halo made of objects of the same mass, figure 5 shows the mean value of the 'excitation time' and the mean number of microlensing events expected for 10^6 stars monitored (in the LMC) during one year as functions of the mass of the deflectors.
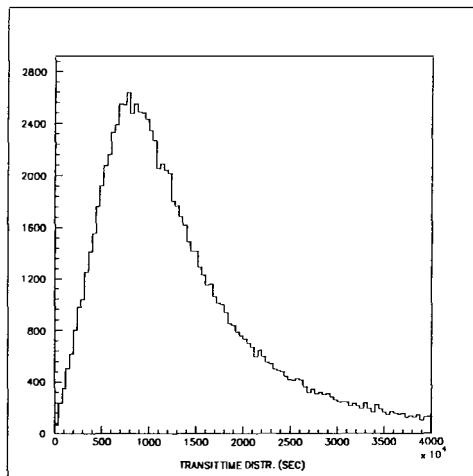
Figure 4 : probability distribution for the "excitation time"
(for 1 solar mass deflectors)
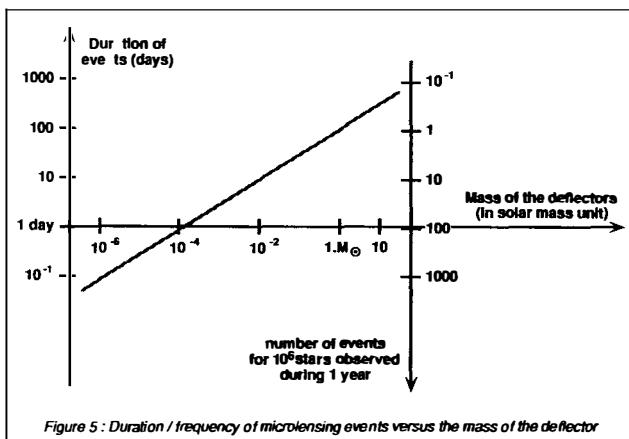
Figure 5 : Duration / frequency of microlensing events versus the mass of the deflector

2.3 Backgrounds and physics summary

The main a priori background to this search is intrinsic variable stars. The requirements of *symmetry*, *achromaticity* and *uniqueness* of the brightening, however, constitute a powerful tool to reject these backgrounds. Also, the magnitude of a star brightened by microlensing should be identical well

before and well after the observed brightening, and the maximum brightening has a very low probability (1.5%) of exceeding 1.50 magnitude. We presently know of no type of variable star that could simulate the signal being searched for [4]. In any case, all candidate stars should be monitored for intrinsic variability following the observation of a brightening that is compatible with a microlensing effect, and the old data available on such stars should also be checked.

To summarize, we have to look for characteristic star brightenings over a wide range of 'excitation' times. In the case of deflectors that are low mass stars (0.1 M$_\odot$) or brown dwarves, the typical excitation time is one month so that a few million stars should be monitored with the highest possible sampling rate (at least every 10 nights) during several months. In the case of 10^{-6} M$_\odot$ deflectors, about 60 000 stars should be monitored at least every 40 minutes during 20 days (see also figure 5). Whereas in the first case, photographic plates are needed to obtain the necessary statistics, the study of the second case requires less statistics but a more frequent sampling of star brightnesses and thus shorter exposure time, so that CCDs are well suited. This mass range for deflectors seems to be the largest reasonable one, because higher mass stars are luminous and would have already been seen (except for possible black holes) and lower mass primordial objects would most likely have evaporated since the time of galaxy formation.

The feasibility of such a project has been studied in detail with existing photographic plates (see below) and leads to the conclusion that it is possible using the European Southern Observatory (ESO) facilities in La Silla, Chile.

3. PHOTOGRAPHIC PLATES

3.1 Feasibility study

Two pictures of the Large Magellanic Cloud (LMC : 30 square degrees) taken 6.5 years apart with the 1 meter Schmidt telescope at La Silla (30x30 cm^2 plates) were analysed in order to study the reproducibility of the luminosity measurement. These pictures were made in similar conditions (1 hour exposure time with a Red filter). They were partially digitized (20 cm^2) in steps of 10 μm by the MAMA measuring machine at the Paris Observatory (MAMA = Machine Automatique à Mesurer pour l'Astronomie). The studied region, not too far from the bar of the LMC, is a quite crowded field (5 to 10 000 starlike objects per cm^2 on the plate).

The algorithm defining stellar objects on these plates that is presently used during the digitizing procedure at the MAMA is simple but rather fast. On the other hand, it is not ideally suited for star finding in regions that are as densely populated as the ones we are interested in, and in the future it will be necessary to develop a more efficient algorithm for the definition of the stellar objects. In spite of this shortcoming, we have tried to match the objects that are identified on the two plates. The comparison between the two plates is performed in a completely automatic way. One plate's coordinates are geometrically adjusted with respect to those of the other using, in a first stage, pattern recognition based on the brightest objects of each plate. In a second stage, a converging fitting procedure is applied, using a large number of objects unambiguously associated after the first stage. After the geometrical adjustment has been done, a large proportion of the objects from the two plates

can be unambiguously considered as images of the same star (applying the coordinates transformation, the two images of a star coincide with a precision better than 7μm). The proportion of stars correctly identified in that way on the two plates is 66%. The remaining (missing) 34% are explained by the already mentioned limitations of the simple star-defining algorithm that we are presently using in connection with the MAMA machine.

Figure 6 shows the distributions of magnitude differences measured in the two plates for objects that are correctly associated (in 4 adjacent intervals in a magnitude range of $\Delta M = 2$). These distributions are strongly peaked at values which are related to the small differences in the recording conditions for the two plates. Their half-widths at half-maximum are less than 0.07 unit of magnitude. For a star in this $\Delta M = 2$ range, the probability to get a magnitude variation larger than 0.32 is 0.2%, which is compatible with the expected rate of variable stars.
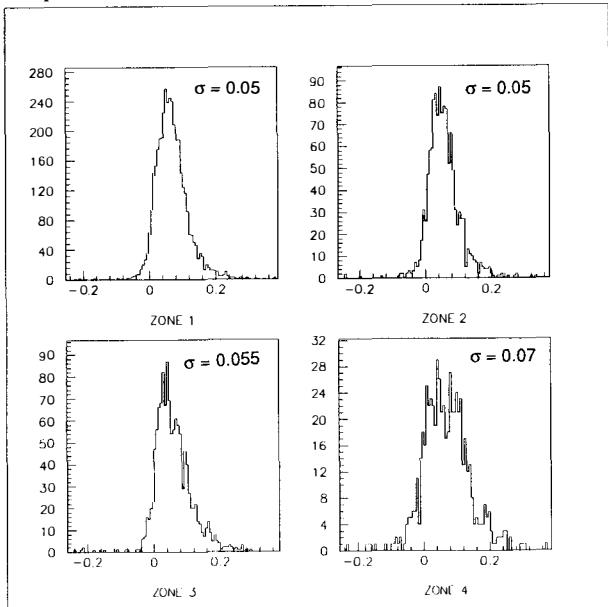
Figure 6 : differences of magnitude measured in two plates for stars
in 4 adjacent intervals in a magnitude range of $\Delta M = 2$

This analysis was performed using the vectorial capabilities of an IBM 3090 machine, which makes the calculations in less than 2 minutes (CPU time) for about 100 000 objects. From the data analysed here, and given the fact that the algorithms used need computing time proportional to the number of objects under study, one can estimate the amount of computing resources that will be needed for a full comparison of two Schmidt plates. The necessary computing time amounts to 1-2 hours on an IBM 3090 and a disk space of about 0.5 Giga-octets is required.

In summary, this feasibility study demonstrates that an automatic program is able to analyse pictures with at least 4000 objects per cm^2 so that the goal of monitoring a few million stars can be reached.

Moreover, we have achieved plate photometry that, in the region where the density-magnitude relationship is nearly linear, is reproducible to 0.07 magnitude, well below the 0.32 magnitude threshold that we have set for the search of microlensing events. Note that (possible) non-gaussian tails in the magnitude variation distribution cannot be a problem because of the large sampling rate of data taking.

3.2 The observing program

The observing program consists in recording about 50 plates taken at the 1 m Schmidt telescope of ESO (La Silla, Chile) ; two plates are needed per night, one with a B (blue) filter and about 30 mn exposure time, the other one with a R (red) or V (visible) filter and 50 mn exposure time. All these plates should be taken during the next period of observation of the LMC (October 1990 to March 1991), with seeing conditions better than 1.5", following a sampling schedule which allows for covering the considered range of 'excitation times'.

This program would allow us to look for microlensing event durations ranging from 3 days to a few months, with at least 5 points measured on the light curve above an amplification by a factor 1.34.

In order to strengthen our conclusions, any candidate star should be monitored for intrinsic variability, either using CCDs (see next section) or photoelectric photometry. We would also compare these plates with already existing ones taken in the last few years and would probably ask for some plates to be taken in the year following this program.

 So far, we have not discovered any background effect that could simulate the signal we are looking for.

4. THE CCD APPROACH

Present CCD performance makes them particularly attractive for photometric observations of a relatively large sample of stars. Their main advantages are a high quantum efficiency, good linearity over a wide dynamic range, an extraordinary stability and reproducibility and an intrinsic digitization. However, the number of pixels remains small with respect to conventional photography (10^9 pixels on a Schmidt plate). An array of CCD's of reasonable price can offer a few million pixels. This, however, is enough to cover the lower part of the mass range (10^{-6} M$_\odot$ to 10^{-4} M$_\odot$).

We propose to build an array of 16 CCD's which will cover an area of 20 cm^2. The CCD's developed by THOMSON-CSF (THX-31157) can be arranged side by side on three edges to minimize the dead zones. Each pixel is 22x22 μm^2 and the total number of pixels is 3.7 10^6. This CCD array can be mounted in the focal plane of the existing GPO telescope (Grand Prisme Objectif) at La Silla, Chile, which has an angular field of view of 2° by 2° and provides pictures of 16x16 cm^2. Many more CCD's than proposed can in fact be accommodated, if necessary, at this telescope. In the first experimental step with 16 CCD's in the focal plane, it will be possible to register some 60 000 stars in an exposure time of about 20 mn. In the focal plane of the GPO, one arc second corresponds to one CCD pixel. This pixel size is of the order of the dispersion given by the seeing conditions which is quite an optimum. The only disadvantages of the GPO are its relatively small aperture (40 cm), its relatively poor equipment (no automatic handling) and the presence of chromatic

aberrations (currently under investigation). With such a telescope and with the 16-CCD array, we would be able to cover the range 10^{-6} M $_\odot$ to 10^{-4} M $_\odot$ for the massive compact objects of the galactic halo.

Obviously, all the software presently developed for the photographic plate analysis can be used for the CCD picture analysis.

5. CONCLUSIONS

The two complementary approaches will cover the range 10^{-1} M $_\odot$ - 10^{-4} M $_\odot$ (using plates) and 10^{-6} M $_\odot$ - 10^{-4} M $_\odot$ (using CCD's) for massive compact objects. Combining our CCD data with another experiment equipped with CCD's looking at the same region of the LMC (from New-Zealand for instance) it would be possible to follow a complete light brightening curve of 12 hours to 2-3 days, corresponding to deflectors of mass around 10^{-3} M $_\odot$ - 10^{-4} M $_\odot$.

We believe that no known variable star could constitute a real background (uniqueness of the occurrence, shape of the excitation curve, achromaticity). In the case of strong indication for a microlensing signal, it would be necessary to compare the study of the stars in LMC with a similar study of an equivalent population of nearby stars, the only difference between these two populations being the optical depth of the halo. The discovery of the microlensing effect on a star, well separated from the phenomenon of variability, would in itself be a beautiful achievement independently of the related question of dark matter.

We believe that the proposed experiment is the most crucial dark matter experiment that can be done with present technologies. The demonstration that the halo is made of massive compact objects rather than a cloud of weakly interacting elementary particles would eliminate one big motivation for low mass scale supersymmetry, destroy an industry of particle theorists calculating relic densities for such particles, and motivate a completely new approach for triggering galaxy formation. On the other hand, the demonstration that MCOs do not exist would provide good evidence that the halo is indeed made of weakly interacting elementary particles.

Finally, while particle theory and cosmology have progressed over the last years through extensive exchanges between theoretical astrophysicists and particle theorists, this experiment will cement a true collaboration between experimental particle physics and astrophysics. We believe that the interplay between particle experimentalists well familiar with the handling of large amounts of data and astronomers well aware of star recognition problems is a unique and invaluable feature of this proposed experiment.

REFERENCES

[1]   See e.g. the following review articles : Faber S.M. and Gallagher J.S., Ann. Rev. Astron. Astrophys., 1979 and Trimble V., Ann. Rev. Astron. Astrophys., 1987

[2]   B. Paczynski, Astrophysical Journal **304** (1986) 1

[3]   See e.g. Primack J.R., Sadoulet B. and Seckel D., Ann. Rev. Nucl. Part. Sci., 1988

[4]   See Petit, 'variable stars',ed. John Wiley

GRAVITATIONAL LENSING
AS A PROBE FOR DARK MATTER

Laurent Nottale

CNRS.
Département d'Astrophysique Extragalactique et de Cosmologie.
Observatoire de Meudon. F-92195 Meudon Cedex. France

ABSTRACT

We show in the present contribution how some statistical effects of lensing may yield unique information on the distribution of matter in the universe and may even probe the relative distribution of dark and visible matter. Such a goal may be reached provided detailed information is known on the observed foreground matter lying in the fields of any of the test sources in a sample of "standard candles".

172

1. INTRODUCTION.

Gravitational lensing is a unique tool for probing the distribution of matter in the universe, and in particular to detect dark matter. This is due to the fact that either mass or surface density directly intervene in the various gravitational lensing effects, and that all levels of the hierarchy of structures, from compact objects and stars to the largest scales may act as lens and as source for lensing.

But clearly, if gravitational lensing may be caused by all mass scales, the observable effects which are expected from them vary with the scale. In the present contribution, we shall consider some of these effects (which had been mainly studied up to now in the frame of a study of lensing itself), from the point of view of the information they may yield about the distribution of matter in the universe. A particular weight will be given here on statistical lensing effects. Neither the analysis of known multiple image configurations nor microlensing are considered here, since they have been the subject of many recent papers and reviews to which we send the interested reader.[1-4]

2. VARIOUS LENSING EFFECTS.

The most common gravitational lens effect is *luminosity magnification* (or attenuation), since it is encountered whatever the physical configuration, provided it corresponds to a perturbation relative to the Friedmann-Robertson-Walker background average universe. To this magnification is always linked a *diameter enlargement*, since Etherington's theorem[5] ensures the conservation of $B(1+z)^4$, where B is the surface brightness and z the redshift, this result holding in any Riemannian universe. It thus applies to various images of a same source (z=constant $\Rightarrow$ same expected surface brightness), and also allows to relate directly the magnitude-redshift relation to the diameter-redshift relation into any universe, whatever inhomogeneous. The diameter-redshift relation may be obtained in the most general frame by solving the Optical Scalar Equations (OSE) which govern the propagation of the cross sectional area of a light beam[3,5-7]. In this equation, the effects of matter and energy (i.e. equivalently, thanks to Einstein's equations, of geometry) on the light beam propagation come from two terms:

(i) A "matter" term, based on the Ricci tensor, which is often interpreted as describing the converging effect of a smooth density of matter included into the light beam. However it plays in fact a far more important role, since a lot of studies have shown that it also yields the effect of a clumpy distribution of matter lying in the light beam [8,9], and the average effect of a clumpy distribution of matter lying outside the beam[3]. Though numerical simulations confirm this last result even for large average amplifications[8,10], unfortunately an analytical demonstration is still lacking in this case, due to the fact that it corresponds to multiple lensing.

(ii) A shear term, based on the complete Riemann tensor and mainly due to individual masses lying outside the beam and to density anisotropy.

It has been shown in the most general way that the luminosity magnification may be written under the form:[3]

$$Amp = \{(1-\kappa)^2 - \gamma^2\}^{-1} \tag{1}$$

where κ and γ reduce to leading order respectively to the matter and shear terms obtained in Einstein linearized theory[8,11]. We recall that the matter term writes typically:

$$\kappa = \frac{4\pi G}{c^2} \, [\int \delta\rho \, dr] \, \frac{c}{_o H}(z_d) \, [1 - \frac{\Delta_d}{\Delta_s}] \tag{2}$$

where $\delta\rho$ is the density difference between the lens and the cosmological background universe and is integrated inside the lens along the line of sight, $L(z_d)$ is a characteristic function of the lens redshift and Δ is the so-called "optical distance"[3,12], which is a monotonic function of redshift . The final term is another expression for the usual distance term written in terms of the angular diameter distances:

$$D_d.D_{ds}/D_s = \frac{c}{H_o} L(z_d) \, [1 - \frac{\Delta_d}{\Delta_s}] \tag{3}$$

The luminosity magnification due to shear by a point mass lens M is defined, in terms of the lens-image impact parameter r , by a γ term which writes:

$$\gamma = \frac{4G}{c^2} \frac{M}{r^2} \frac{c}{H_o} L(z_d) \, [1 - \frac{\Delta_d}{\Delta_s}] \tag{4}$$

For very large scales (superclusters, filaments, sheets of matter and most clusters of galaxies), "Ricci" luminosity (and diameter) magnification remain the only lensing effect. Not being related to a morphological, spectral or variability signature, it can be identified only by statistical methods. Two of them have been particularly considered: Perturbation of the Hubble diagram of distant sources by foreground matter,[13,14] and change in the number surface density of galaxies or quasars,[3,10,15] with the corresponding selection effects which are implied by and related to these phenomena.

3. PERTURBATION OF HUBBLE DIAGRAM BY FOREGROUND MATTER

It has been suggested several years ago by Karoji and Nottale[13] that rich clusters of galaxies magnify in a statistically significant way the luminosity of background sources lying behind them. Their method consisted in dividing a sample of standard candles into two sub-samples, those which lie behind a

foreground cluster (B) and those whose light has not crossed any inhomogeneity on its way to the observer (A), then in comparing the relative distribution in the Hubble diagram of subsamples (A) and (B). The so-called "Hubble modulus":

$$h_m = log\{cq_0^{-1}z+cq_0^{-2}(q_0-1)[-1+(1+2q_0z)^{1/2}]\} - 0.2m \qquad (5)$$

is used for such a comparison, since it characterizes the deviation from the mean Hubble line.

The effect was later confirmed on a sample of brightest cluster galaxies by a direct comparison with the prediction from gravitational lensing expectation.[14] Each cluster in the sample was placed relatively to foreground Zwicky clusters in the same field, so that the distance term $D_d.D_{ds}/D_s$ (Eq. 3) could be computed for each configuration. From the observed impact parameters the density term $\int \delta\rho.dr$ was also estimated (to a multiplicative constant depending on the absolute density of lensing clusters, which was assumed constant as a first approximation), so that the matter term κ could be estimated for each relative configuration (foreground cluster/background cluster).

Note that most of the time the redshifts (or distance classes) of foreground and background objects are different enough to ensure that they actually do not lie at the same distance, but that anyway the form of the distance term (Eq.3) which vanishes when the lens comes close either of observer or of source, ensures the vanishing of the contribution of such configurations to statistics from the point of view of lensing.

Another remark on which to stress here is about the validity of Zwicky clusters for leading such a study. While the Zwicky catalog may indeed be criticized if one is interested in the identification and analysis of individual clusters, since it is based on counts made up to a given limiting magnitude and on number surface density excess relative to the field, this method conversely implies that it is well adapted to statistical studying lensing effects, in particular by large scale structures. This is because it is surface density which intervenes in the intrinsic term of the lensing formula, so that the same global effect will be expected from several superposed small clusters (lying at comparable distances) or from one rich cluster having the same integrated density.

A third point to be made is that samples of brightest cluster galaxies, and more generally the Hubble diagram of "standard candles", have been given up as useful tool for cosmology, after the chess of the various attempts to get from it the value of q_0. However it is important to notice that the Hubble diagram keeps all its value for statistical lensing studies. Indeed the rejection came from the fact that evolution effects were finally understood to be far larger than the searched effects of q_0, thus preventing from its measurement. So the question here is the relation of

evolutionary effects and q_o effect to the statistical effect of magnification by lensing. First, while luminosity evolution depends only on the source redshift, it is one of the significant signature of lensing that it depends also of lens redshift and lens density. No other physical phenomenon may lead to correlations between what is thought to be intrinsic to a source (its "absolute" luminosity, when lensing is not accounted for) and the properties of foreground matter which may lie at cosmological distances from the source. Second the effect of q_o is also fixed for a given redshift, while a dispersion is introduced in the lensing effect for given source redshift by the various configurations which may be encountered (lens redshift, density and angular distance).

The final result was a significant correlation between the deviation from mean Hubble line (Eq.5) and the expectation from lensing magnification (Eq.2).[14] Unfortunately, though the existence of the effect itself was established independantly from the initially assumed value of q_o (or σ_o), several uncertainties prevented from being able to state precisely at this stage on the amount of dark matter required for it. In fact information may be gained by this method on both q_o and σ_o, since q_o intervenes in the Mattig relation, while the unknown density term in the lensing formula (which is given by the slope of the regression line) is directly proportional to σ_o. The results are as follows:

*The high value of q_o found for the whole sample (uncorrected for lensing), $q_o \approx 1.6$, falls down to $q_o < \approx 0.7$ when only the "isolated" (i.e. unaffected by lensing) subsample is taken into account. This is a first indication that a selection effect is at work, since an all sky sample randomly distributed with respect to foreground matter should not have its Mattig relation changed in the mean because of energy conservation.

*The slope of the regression line (between the observed and expected deviations from Hubble line) corresponds to an apparent value of $\Omega_o > \approx 1$. However this result should not be taken for granted, since it can be shown that the slope is over-estimated because of a selection effect. Indeed the KSW brightest cluster galaxies lie in regions of the sky where the number density of foreground Zwicky clusters is 2 to 2.5 times larger than for the average catalog and where the Zwicky clusters are significantly more compact than in the average sky.

Finally, though our results are compatible with $\Omega_o > \approx 1$, such an interpretation does not take into account the selection effect which has been actually evidenced independantly. On the contrary the choice $q_o = \sigma_o < \approx 0.2$ combined with the expected bias consisting in a preferential selection (precisely due to luminosity magnification by lensing) of clusters situated behind the richest foreground clusters, allows a complete interpretation of the whole set of data. We conclude that the method should now be extended to other samples of galaxies and clusters, which we intend to do,[16] and that it will become really efficient as a probe for the density

of the universe only once one succeeds in quantifying precisely the selection effect by gravitational magnifications.

4. PROBING THE DISTRIBUTION OF DARK MATTER FROM LENSING.

One of the main problems of present observational cosmology is the precise distribution of dark matter. We are now convinced that indeed most of the mass (more than 90%, i.e. $\Omega_o \approx 0.1$, maybe 99%, i.e. $\Omega_o \approx 1$) of the Universe is dark, but we still dont know whether it is distributed like the visible matter or in a completely different way. Answering this question would be relevant in the understanding of the nature of the main constituents of the Universe. In particular the extreme case of a completely uniform density of WIMPs with $\Omega \approx 1$ added on $\Omega \approx 0.1$ baryonic matter distributed as light is, would be extremely difficult to detect.

We propose herafter a test of such a distribution using the comparison between dynamical measurements and lensing measurements of masses. It is based on the theorem of energy conservation for gravitational magnifications. It was established globally by Weinberg[17]: On the whole 4π sphere, amplifications are balanced by attenuations such that the Mattig relation is correct in the mean, and then locally by Nottale[3]: the same result holds into any cone bounded by Friedmann-Robertson-Walker null geodesics. To lowest order, this happens for cones resting upon domains of the Universe into which the total mass equals that which would have been found in a totally homogeneous Friedmann model of same average density.

This result comes from the fact that the part of the matter term which is intrinsic to the lens writes $\int \delta\rho.dr$ in the hereabove formula for κ (Eq. 1 and 2), i.e. one should integrate not over the density in the lens, but over its departure from the mean cosmological density, $\delta\rho = \rho - \rho_o(1+z)^3$. On the contrary the measurement of masses by dynamical methods is a local one and will not take into account a possible very large scale component showing no density gradient at the scale of the lens. In fact we find that both measurements would be related, for a mass M_{dyn} contained into a radius r, by the formula:

$$4\pi \int \delta\rho.dr = 6 \frac{M_{dyn}}{r^2} [1-(\rho_o/\rho)^{2/3}]. \qquad (6)$$

This means that when the dynamical method measures M_{dyn}, the lensing method measures in fact $M_{dyn}[1-(\rho_o/\rho)^{2/3}]$. Assume now that we compare the average luminosity of sources situated behind such a lens (B) to reference sources (A) whose light has travelled in the average background universe on its way to the observer (following the Karoji-Nottale[13] method) and that we find that they are

magnified by a given amount: $\{log(cz)-0.2m\}_B > \{log(cz)-0.2m\}_A$. From this magnification one may deduce the mass of the lens.

When computing this mass, most authors use the uncorrected (thus non energy conserving) formula. It is clear that indeed the corrective term $[1-(\rho_o/\rho)^{2/3}]$ plays no role when the lens is a galaxy or a rich cluster of galaxy, for which $\rho_{light}>10^3\rho_{olight}$. Whatever the value of Ω_o, even if it is as large as 1, one gets in these cases $\rho_o/\rho\ll1$ and the usual uncorrected formula holds.

Now if the lens is a supercluster or an edge-on filament or sheet of galaxies, one may still obtain detectable amplifications thanks to the large distances over which the density is integrated; it has been shown[3] that the thin lens formula still holds in this case. But now the eventual existence of a very large scale component of dark matter would play an important role. Assume for example that the density of visible matter in the lens is found to be 10 times the mean visible matter density of the Universe. An additional very large scale flat distribution of dark matter corresponding to $\Omega_o\approx1$ would imply that $\rho/\rho_o\approx2$ instead of 10, and the use of the uncorrected formula would yield a discrepancy by a factor of ≈3 between the dynamical and lensing masses, while they would agree if $\Omega_o\approx0.1$.

Another way to look at the test is to remark that the corrective term may become so important, that while a significant detection of a statistical effect of magnification by large scale structures would be expected from their observed density if dark matter is distributed as light is, it would become unobservable for large Ω and strong bias. So the fact that such effects do exist in a statistically significant way (on brightest cluster galaxies[14] and absorption line quasars[3]) goes in the direction of the exclusion of a strongly biased distribution of dark matter, in particular in the case $\Omega_o\approx1$. But this case would also encounter strong difficulties in the hypothesis of light exactly tracing matter.

Indeed it has been remarked already several years ago[12,18] that, under this last hypothesis, even in the case $\Omega_o\approx0.1$ the richest Coma-like clusters of galaxies may produce, when placed at optimal distance for lensing (i.e. $z_d\approx0.3-1$, which yield a maximal value for $L(z_d)\approx0.2$) a matter κ term which reaches the critical value $\kappa=1$ i.e. very large magnifications. For example one gets for an isothermal lens:

$$\frac{4\pi G}{cH_o} \left[\int \delta\rho \, dr\right] = \left[\frac{M}{5 \; 10^{15} \; M_\Theta}\right] \left[\frac{1 \; Mpc}{r_p}\right] \tag{7}$$

where M is the total cluster mass included into 3 Mpc radius and r_p is the line of sight impact parameter. With a 200 kpc impact parameter and maximal $L_d\approx0.2$, one gets $\kappa\approx M/10^{15} \; M_\Theta$, while the total mass of Coma, assuming light tracing matter and $\Omega_o\approx1$ is $M\approx10^{16}M_\Theta$ (from an average density in 3 Mpc radius estimated to be 2000

times the cosmological density). So even with $\Omega_o \approx 0.2$, one would get a magnification of 1 mag. behind such a cluster, thus significantly perturbing the Hubble diagram of more distant objects. Then we argue that if $\Omega_o \approx 1$, such an effect could be obtained from lensing clusters ≈ 5 times less rich than Coma (i.e. Virgo-like), which seems to be excluded from the present status of observations.

5. CONCLUSION

We have been concerned here with the possible use of some statistical gravitational magnification effects in order to probe the distribution of dark matter in the Universe. The present status of data does not allow yet to reach a firm conclusion. However we believe that these methods may become in a near future very good tools for this aim, once progress are made concerning:

(i) The quantification of selection effects due to lensing. This implies in particular a better theoretical understanding of the statistics of lensing by all scales of the hierarchy of structures.

(ii) The increase of data, concerning in particular the observed distribution of foreground matter around any source of a test sample. Such information is now in progress, for example for a sample of very distant galaxies[19], and is also relevant to point i).

REFERENCES.

1. Canizares, C.R., 1987, in "Observational Cosmology" , *IAU Symp. 124*, eds. Hewitt, Burbidge, Zang (Dordrecht: D. Reidel), p. 729
2. Turner, E.L., in "Dark Matter", *8th Moriond Astrophys. Meeting*, ed. Audouze & Tran Thanh Van, Frontières, p. 347
3. Nottale, L., 1988, *Ann. Phys. Fr.*, **13**, 223
4. Nottale, L., 1988, in "Dark Matter", *8th Moriond Astrophys. Meeting*, ed. Audouze & Tran Thanh Van, Frontières, p. 339
5. Etherington, I.M.M., 1933, *Phil. Mag.*, **15**, 761
6. Sachs, P.K., 1961, *Proc. Roy. Soc. London*, **A264**, 309
7. Kantowski, R., 1969, *Astrophys. J.*, **155**, 89
8. Young, P., 1981, *Astrophys. J.*, **244**, 756
9. Dyer, C.C., Roeder, R.C., 1981, *G.R.G.* , **12**, 1157
10. Peacock, J.A., 1986, *Mon. Not. Roy. Soc.*, **223**, 113
11. Bourassa, R.R., Kantowski, R., 1975, *Astrophys. J.*, **195**, 13
12. Nottale, L., Hammer, F., 1984, *Astron. Astrophys.*, **141**, 144
13. Karoji, H., Nottale, L., 1976, *Nature*, **259**, 31
14. Hammer, F., Nottale, L., 1986, *Astron. Astrophys.*, **167**, 1
15. Canizares,C.R., 1981, *Nature*, **291**, 620
16. Nottale, L., Marchandon, S., in preparation
17. Weinberg, S., 1976, *Astrophys. J. Lett.*, **208**, L1
18. Turner, E.L., Ostriker, J.P., Gott III, J.R., 1984, *Astrophys. J.*, **284**, 1
19. Hammer, F., Le Fèvre, O., 1990, Astrophys. J., in press

ARCS AND ARCLETS
IN CLUSTERS OF GALAXIES

Y. Mellier and P.-Y. Longaretti

Observatoire Midi–Pyrénées

14, avenue E. Belin

31400 Toulouse, France

ABSTRACT

Arcs and mini–arcs (hereafter "arclets") are gravitationaly distorted images of high-redshift background galaxies, lensed by a rich foreground cluster of galaxies. The study of these distorted objects should give informations on the evolution of galaxies, the dark matter distribution in large scale structures and the cosmological parameters. The observation of these very faint objects is extremely difficult but first results have already been obtained on the richest clusters of galaxies. In 50% of the cases, arcs and/or arclets have been detected leading to the measurement of some new arc redshifts. The analysis of A1689 by Tyson confirms how powerful arclets can be for the understanding of the dark matter distribution in clusters and is very encouraging for the continuation of the program of arc survey in clusters of galaxies we started a year ago.

Observing the arcs : introduction

During the last five years, two major discoveries opened a new field of investigation in observational cosmology with particular implications for the understanding of the dark matter distribution in clusters of galaxies, the large scale structures of the universe and related cosmological problems. The first one is the observation of arcs and arclets which were subsequently identified as images of background galaxies gravitationally distorted by rich clusters[1,2,3], and the second is the discovery by Tyson of a rich population of faint blue galaxies, whose redshift is still unknown but probably larger than 1[4].

Arcs and arclets : Soucail et al. [1] and Lynds and Petrosian[2] first stressed the strange shape of the unknown structures observed in a few distant clusters of galaxies. Finally Soucail et al.[3] identified them as images of lensed background galaxies by measuring the redshift of the giant arc in A370 (z=0.724), twice as large as the cluster's (0.374). More recently, new arc redshifts were obtained[5,6], all of them larger than the cluster's, which confirms their gravitational origin.

In a second step, Fort et al.[7] announced the discovery of very faint blue extended objects in A370. These structures are not as elongated and not as bright as the giant arc and their very low surface brightness makes the measurement of their redshifts quite hopeless (in the near future) with present-day instrumentation. However, the comparison of their color index with model predictions leads to the reasonable conclusion that they are located at redshifts larger than 1. Fort et al. concluded that their distorted shape could be a gravitional effect of the cluster-lens.

The blue Tyson's population : in 1988, Tyson[4] mentioned the discovery of a high density of very faint blue galaxies which roughly fill most of his ultra-deep CCD frames obtained in empty fields. Tyson excluded the possiblity that these objects are nearby dwarf galaxies and concluded that they are probably galaxies at redshifts between 1 and 3 uniformly filling the sky at a magnitude $B_J = 28$.

Although these two discoveries were made independently, it rapidly appeared that they were certainly related. This can easily be understood as follows : if we assume that Tyson's population is at redshifts larger than 1 and that a rich cluster of galaxies is on the line of sight, one can predict that the cluster-lens should distort the background galaxies, leading to the formation of a few number of giant arcs and numerous arclets. We simulated this configuration in the figure. In this example we randomly paved the sky with a single redshift distribution of galaxies at z=1.2 and a projected density given by Tyson at a limiting magnitude of B_J=28. The typical value of the field of view (2'x 2') is the expected one on 4m-class telescopes with ordinary CCDs and the cluster is modeled using standard parameters $(\sigma_{los} = 1000 \text{ km/s}, r_c = 200 \text{ kpc}, z = 0.23)$. The similarity of the resulting image with what is observed in A2218 or A2390 is amazing and suggests several comments.

DISTORTION GRID
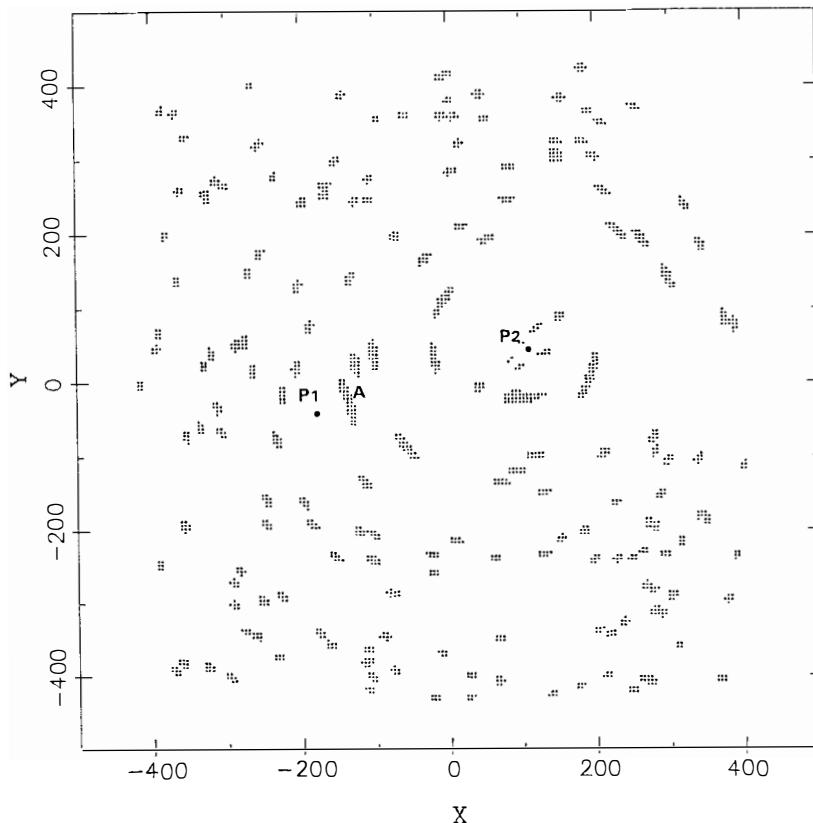
Simulated distortion grid : the sources are assumed circular and randomly distributed at a redshift of 1.2. The cluster–lens is at a redshift of 0.23 with standard structural and dynamical parameters as expected for a rich cluster of galaxies. The cluster is composed of two potential wells located at P1 and P2. The opposite curvature of the arc A is a direct proof that this cluster is bimodal.

First, one can expect that a scientific program based on ultra-deep CCD-imaging of distant clusters of galaxies will yield a wealth of new arclets, since the clusters must act like gravitational lenses on Tyson's population. Second, this large number of arclets per cluster should efficiently probe the mass distribution of the cluster itself. The simulations show that the distorted images are not randomly distributed but closely follow the general shape of the projected mass density. This is rather clearly demonstrated on the figure presented here, where the bimodal shape of the cluster is recognized throughout the distribution of the arclets.

Deep CCD-imaging and numerical modeling are the tools we need for such a program. However, some major questions immediately arise : what kind of astrophysical information can we expect from the observation of arclets? Moreover, the surface brightness of the faint arclets is about $28-29$ arcsec^{-2}, and one can question the feasability of the observations and the methodology of the data reduction and processing techniques. In what follows, we try to address these points which are now our main concern in the framework of an ESO Key-Program.

Observing the arcs : why?

The study of arcs and arclets in clusters of galaxies can give informations on the lens (rich clusters of galaxies), the sources (high redshift background galaxies) and Cosmology (fundamental cosmological parameters).

The lens : in the case of small deviations, the deflection angle of a transparent, spherically symetric and stationary lens is given by $\alpha = 4G/c^2 \ M(b)/b$ where b is the impact parameter and $M(r)$ the total masse contained within the radius r. In principle, it should be possible to infer the potential well of the deflector from the position and curvature of giant arcs. However, models suffer from a too large number of unconstrained parameters and in practice numerous models can reproduce a single giant arc, as in the case of A370. More interesting results can be expected when many arcs and arclets are observed; from an analysis of the image distortions, it should be possible to model the projected distribution of mass within the cluster. Furthermore, if a high density of arclets is observed, we can achieve a real mapping of the mass distribution with a typical spatial resolution of 10", much better than the capability of present-day X-ray satellites.

The sources : giant arcs are extremely amplified images of background galaxies. Such a strong amplification occurs for sources located near a critical line of the potential, where the merging of different images can take place. Due to these large amplification factors, and by co-adding each pixel of an arc, we have an opportunity to get high signal-to-noise spectra of very distant galaxies and compare them with models of galaxy evolution. As these sources are ordinary field galaxies, this approach has the advantage of studying an unbiased sample of distant objects. Unfortunately,

the redshifts and spectral energy distribution of 3 arcs only have been measured to date (A370 : $z_s = 0.724$, A2390 : $z_s = 0.913$, Cl2244 : $z_s = 2.35$). However, using together spectroscopic data and infrared photometry, first preliminary constraints on galaxy evolution have already been obtained[8].

In addition, the detailed observations of microstructures in the arcs can constrain the shape of the source. In such a case, high resolution imaging, good photometry, spectroscopy and modeling can be used simultaneously to infer the morphology of the background galaxy.

Finally, it should be stressed that the number of arclets observed in rich clusters of galaxies depends on the redshift distribution of background sources. Therefore, the occurence of arclets is a direct test of the Tyson's conjecture that his very faint blue galaxies really are at redshifts larger than 1.

Cosmology : in the case of an isothermal sphere, the deflection angle can be expressed as $\alpha = 4\pi \ (\sigma_{los}/c)^2 \ D_{ls}/D_{os}$ where σ_{los} is the line of sight velocity dispersion, D_{ls} and D_{os} are the angular distance from the lens to the source and from the observer to the source respectively :

$$D_{ij} = 2/H_0 \frac{[(1 - \Omega_0)(G_i - G_j) + (G_i G_j^2 - G_i^2 G_j)]}{\Omega_0^2 (1 - z_i)(1 - z_j^2)}$$

with $z_i < z_j$ and $G_i = \sqrt{1 - \Omega_0 z_i}$

Therefore, α is independent of H_0 and in principle, one can compute the value of Ω_0.

However, this technique is strongly model-dependent and in fact quite hopeless at present. Most of the observational problems related to the estimation of the one-dimensional velocity dispersion can be bypassed if at least two arcs are detected since the ratio α_1/α_2 does not depend on this quantity. But the assumption of an isothermal sphere is rather crude, and strong constraints on the structure of the potential are needed before we can measure Ω_0 with any confidence. Such constraints might be obtained in a statistical way if a large number of arclets (one hundred or so) are observed in the cluster.

Observing the arcs : where?

A priori, any cluster of galaxies could be a good lens candidate for the detection of arcs and arclets. Indeed, all the distorted galaxies already observed are found in the richest clusters. Although this could be interpreted as a selection bias (only the richest distant clusters of galaxies are observed), it is obvious that the very rich clusters will

be the best candidates : if we assume that Tyson's distant blue galaxies are uniformly distributed on the sky-plane, the number of arclet candidates (galaxies which can be lensed) depends on the surface defined by the critical radius : $N_{arclet} \propto R_c^2$. In the case of an isothermal sphere, $R_c = 4\pi \ (\sigma_{los}^2/c^2) \ D_{ls}/D_{os}$. Therefore $N \propto \sigma_{los}^4$! This strong dependence on the velocity dispersion shows that only the very rich and dense clusters will be excellent candidates. The strategy for cluster selection is then dictated by the following criteria :

$N_a > 100$ and $\sigma_{los} > 1000$ km/s and $L_x > 10^{45}$ erg/sec

In addition : $0.15 < z_{lens} < 0.8$

In fact, not all these data are available but for a very small number of clusters and in order to get a significant sample we generally relaxed the contraints and pre-selected most of the clusters which satisfy two of these criteria. In such a case we have built a cluster-lens catalog of about 60 clusters in both hemispheres.

Observing the arcs : how?

Although presently about ten of the observerd giant arcs are rather bright structures, most of the arclets have extremely low surface brightness and their detection requires special observational techniques and very long integration time.

The signal-to-noise ratio depends on the integration time t as follows :

$$[S/N] = \sqrt{t} \ \sqrt{C \ S_t \ \varepsilon \ s_p} \ 10^{[-0.4(\mu_b - \mu_s)]}$$

with :

$S_t =$ telescope surface

$\epsilon =$ total efficiency of the instrument

$s_p =$ pixel area in arcsec^2

$\mu_b =$ surface brightness of the object

$\mu_s =$ surface brightness of the sky

$C =$ calibration contant

We have neglected the readout noise of the CCD.

For instance, using a low noise CCD with a 30μ pixel-size at the Prime Focus of CFHT (F/3.7) the detection limit for an object of surface brightness equal to 28 arcsec^{-2} (S/N=3) is reached in about 5 or 6 hours!

We can now summarize the technical requirements as follows : 4m telescopes, low noise CCD and shift–and–add techniques during exposure to ensure confident detection and accurate photometry. The same technique must be applied on a comparison field since we try to detect an excess of distorted object in the cluster field. Therefore no more than 2 clusters can be done completely during an observing run. The cluster candidates have to be very good ones!

Observing the arcs : so what?

The various aspects described above form the starting points of a common program of arc survey in clusters of galaxies in collaboration with other institutions in the USA, the UK and Spain. Final results of such a long term program cannot be expected before two years and no significant statistics about the occurence of arcs and arclets can be given, mostly because the observed sample is too small but also because the clusters have not been observed at the same limiting magnitude. However, more than 10 rich clusters have a giant arc near their core[9] and about 50% of our sample for which deep detection have been undertaken shows arcs or arclets.

The most interesting results were obtained by Tyson in A1689[10]. In this cluster more than 60 arclets have been detected. The radial distribution of the distortion of arclets is null at the origin, slowly increases with radius before reaching a maximum value at a typical radius R_c and decreases towards a constant value at infinity. This behaviour is expected if the potential is isothermal with critical radius R_c. The distortion is given by :

$$D = D_0 \; \frac{\kappa^2 + \mu^2}{\kappa^2 - \mu^2}$$

with κ = convergence and μ = shear.

For an isothermal sphere $\kappa^2 = (1 - (R_c/2r)^2)$ and $\mu^2 = (R_c/2r)^2$.

Therefore :

$$D = D_0 \; \left(1 + \frac{1}{2} \; \frac{R_c^2}{r(r - R_c)}\right)$$

This expression nicely reproduces the shape of the observed data in A1689; this strengthens our hopes that arclet detection can be a new diagnosis tool for the mass distribution in clusters of galaxies. The similarity between Tyson's observation and the distortion distribution expected for an isothermal sphere does not guarantee that A1689 is an isothermal sphere. However, it qualitatively agrees with the X-ray map

which shows for this cluster a rather smooth, circular and a centrally peaked emission pattern .

For the time being, A1689 is the only cluster of galaxies for which ultra-deep CCD imaging in two bands is now finished. In the case of A370, although one giant arc is detected as well as 6 small arcs or arclets, the photometry is not so deep and the diagnosis about this cluster is questionable. The analysis will probably be more difficult since it has an elliptic shape with possibly two density peaks located on the brightest cluster members. The best candidates which are now our priority are A370 and A2390 for which we do observe arclets and we have already measured the redshifts of their giant arcs. The observation of a giant arc in addition to a lot of arclets could be essential to properly recover the potential.

Observing the arcs : concluding remarks

In this paper, we reviewed most of the questions related to the realisation of an "arc survey in clusters of galaxies". All the observational and data reduction problems are now understood and solved, except bad weather conditions! However, the ultimate step remains the determination of the potential well of the cluster and the related problem of the dark matter distribution. First attempts to model clusters with giant arcs lead to quantitative results on the shape of the projected mass density and the mass–to–light ratio. However, the number of model parameters is still too large and several different solutions are possible using the same observational data (as for the giant arc in A370). The most promising approach is probably a statistical analysis of the distribution of arclets in these clusters. Arclets are numerous, well distributed everywhere in the cluster field with a good spatial resolution. Their shape (distortion) and position are strongly dependent on the mass distribution within the lens. This is the reason why we are now working on the determination of this mass distribution based on arclet statistics. In a first step we are studying the simplified case of a spherically symmetric cluster. The basic questions are : is it possible to determine in a unique way the projected potential from the arclets distribution and properties? And if so, how many arclets do we need to find the mass distribution at a good confidence level?

The observation of arclets in distant clusters of galaxies suffers from the limiting capability of present-day detectors and telescopes. It is obvious that new telescopes such as the HST, ROSAT or the VLT are now the most promising tools for the realisation of the program of arc survey in clusters and the mapping of the dark matter distribution in large scale structures.

The "arc survey in clusters of galaxies" is an international collaboration between the Observatoire Midi–Pyrénées (B. Fort, J.–F. Leborgne, P.–Y. Longaretti, G. Mathez, Y. Mellier, J.–P. Picat, G. Soucail), the Bell Laboratories (A. J. Tyson,

P. Guhathakurta), Princeton (E.L. Turner) the Universitat de Barcelona (R. Pello, Y. Sanahuja) and the University of Durham (A. Aragon, R.S. Ellis).

References :

[1] Soucail, G., Fort, B., Mellier, Y., Picat, J.–P., 1987 : *Astron. Astrophys.*, **172**, L14

[2] Lynds, R., Petrosian, V., 1986 : *Bull. Am. Astron. Soc.*, **18**, 1014

[3] Soucail, G., Mellier, Y., Fort, B., Mathez, G., Cailloux, M., 1988 : *Astron. Astrophys.*, **191**, L19

[4] Tyson, J.A., 1988 : *Astron. J.*, **96**, 1

[5] Soucail, G., Mellier, Y., Fort, B., Mathez, G., Cailloux, M., 1990 : *in* "Gravitational Lensing". Toulouse, September 13–15 1989. Y. Mellier, G. Soucail, B. Fort Eds.; Springer Verlag 1990.

[6] Pello, R., Le Borgne, J.–F., Mathez, G., Mellier, Y., Sanahuja, B., Soucail, G., 1990 : in "Gravitational Lensing". Toulouse, September 13–15 1989. Y. Mellier, G. Soucail, B. Fort Eds.; Springer Verlag 1990.

[7] Fort, B., Prieur, J.–L., Mathez, G., Mellier, Y., Soucail, G., 1988 : *Astron. Astrophys.*, **200**, L5

[8] Aragon, A., Ellis, R.S., 1990 : in "Gravitational Lensing". Toulouse, September 13–15 1989. Y. Mellier, G. Soucail, B. Fort Eds.; Springer Verlag 1990.

[9] Fort, B., 1990 : "12th Conference on General Relativity and Gravitation", *proceedings.* Boulder, Colorado, july 2–8 1989.

[10] Tyson, J.A., Valdes, F., Wenk, R.A., 1990 : *Astrophys. J.*, **349**, L1

ARCS IN CLUSTERS OF GALAXIES:
DISTRIBUTION OF DARK MATTER AND STATISTICS

Vahé Petrosian

D.A.E.C., Observatoire de Meudon and
Center for Space Science and Astrophysics

and

Anton G. Bergmann

Center for Space Science and Astrophysics
Stanford University, Stanford, California 94305
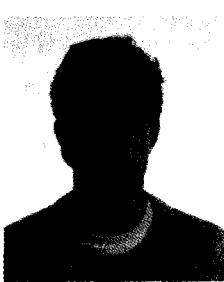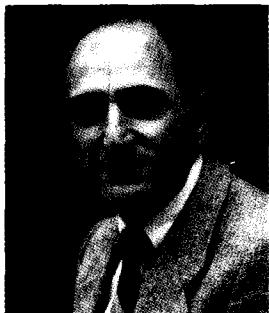
ABSTRACT

We summarize here briefly the properties of the Giant Luminous Arcs in two clusters of galaxies and the inferred distribution of the dark matter in these clusters. We refer to earlier publications for details. We also summarize results from some preliminary work on the reconstruction of the source producing the arcs. Finally, we comment on statistics of the arcs and we discuss the expected relative probabilities of occurrence of arcs with various sizes and shapes and the implication of these for observations.

I. INTRODUCTION

Of the three candidates among the sample of about forty clusters of galaxies observed by Lynds, Petrosian, and Sandage (circa 1976), the giant arc in one (Abell 370 at a redshift $z = 0.373$) is definitely a gravitational lens image of a more distant blue galaxy with redshift $z = 0.725$ [1,2]. Two different spectra of the other well defined arc in Cl2244-02 (redshift $z = 0.328$) show a single emission line at two different frequencies implying widely different redshifts, $z = 0.83$ [1] and $z = 2.2$ [3]. Even though both these redshifts are uncertain, it is unlikely that this arc is not a gravitational lens image of a more distant galaxy. No spectral observation of the little arclets in the third cluster (Abell 2218) is known.

Among the various other arcs or arclets discovered subsequently, the redshift of the one in Abell 2390 is 0.913, which again is larger than $z = 0.23$ of the cluster [4]. A list of the clusters containing arcs and arclets can be found in the proceedings of the Toulouse Workshop [5]. Assuming that all of these objects are gravitational lens images, we have a sizable body of data to begin to study the phenomenon and to make deductions about the sources, lenses, and cosmological structures. It is clear that a great deal can be learned about the relative distribution of light and dark matter in clusters and about the nature of high redshift galaxies. We will summarize the results obtained thus far in Section II. The details of these can be found in various journal articles and conference proceedings, especially the two on gravitational lensing [5,6]. In Section III we discuss a third but not so well understood aspect of the arc phenomenon, namely the absolute and relative probabilities of finding arcs with different lengths and magnifications.

The number of images known at the present time is too small for definite conclusions about the statistics. Furthermore, there is considerable bias in the selection of clusters, especially for the observations subsequent to the initial discovery of the two giant arcs. The sample of clusters among which the first three cases were discovered is not an ideal sample for the task of determining absolute and relative probabilities, but nevertheless is somewhat unbiased because it was selected for some other purpose. It is estimated that

about one in ten of the clusters in this sample showed arcs or arclets [1]. This is in conflict with the simple estimate of one in hundred (or less) for having a source from a population as numerous as galaxies align within one arcsecond of a cluster center. Such an alignment is required for both the Abell 370 and Cl2244 arcs [7,8]. Even more troublesome is that simple arguments indicate that for every arc with magnification greater than 20 (length $20''$ and width $1''$) one would expect about four with magnification greater than 10 and around sixteen with magnification greater than 5. This is not what is seen. Among the original sample we see two with magnification of about 20 and in one case several arclets of magnification less than three.

II. DARK MATTER DISTRIBUTION AND SOURCE STRUCTURE

Most of the results on these aspects of the problem by us or others have been published. Consequently, here we only give a brief summary of the results primarily from our work [8] and those from other publications [9,10,11] which have attempted a detailed modeling of the two well delineated arcs in Abell 370 and Cl2244.

A) Distribution of Mass in Clusters

The following conclusions, listed in order of their degree of certainty, are obtained from the construction of images that match the shape and size of the observed two arcs.

1. It is clear that one can state unequivocally that there is dark matter in the cluster in the sense that the mass to blue light ratio $M/L_B \geq 200 M_\odot/L_\odot h_{50}$ in the central parts of these clusters (i.e. within the radius of the arcs). Here h_{50} is the Hubble constant in units of 50 km s^{-1}Mpc^{-1}. This relation for the mass to light ratio can be obtained from the fact that for a spherically symmetric lens at the radius of the Einstein Ring r_e (the critical line of the lens mapping at which the tangential magnification is infinite) the mean surface density $\overline{\Sigma} = M(r_e)/\pi r_e^2$ must be equal to the critical density $\Sigma_c = c^2 D_s/(4\pi G D_l D_{sl})$, where $M(r)$ is the mass within the radius r and D_s, D_l, and D_{sl} are the angular diameter distances to the source, the lens, and between the source and the lens. The critical density Σ_c describes the redshift and cosmological dependence

of the models and varies by less than twenty percent for cosmological models with matter and vacuum density parameters in the range $0 < \Omega_{mat} < 2$, $0 < \Omega_{vac} < 1$, and by less than fifty percent for source redshifts in the range $0.7 < z_s < 2$ (note that the lens redshifts z_l are well known). The total mass to light ratio, however, could be much higher depending on the distribution of the dark matter outside the radius of the arcs. This is why in some cases [8,9,11] much higher masses are derived.

2. The dark matter cannot be concentrated within or immediately around the galaxies in Abell 370 and must be distributed throughout the cluster [8,9,11].

3. For distributions with finite central surface density Σ_o, effective radius r_o, and total mass $M \propto \Sigma_o r_o^2$, the effective radius r_o cannot be much larger than the radius of the arc, r_{arc}, otherwise M and therefore M/L_B become excessive [8] and $\Sigma_o \to \Sigma_c$. The latter condition would have to be attributed to a coincidence. Furthermore, for larger radius r_o the lens can be approximated by an infinite uniform sheet which produces large radial magnification (thick arcs). This means that arcs with thickness less than one arcsecond (as is observed after correction for seeing) can be produced only by a smaller (therefore higher surface brightness) and more closely aligned source. All these factors decrease the theoretical probability of obtaining highly magnified arcs which as mentioned above is already lower than that observed.

4. For a singular isothermal gas sphere models (or for nonsingular models with a core radius $r_o << r_{arc}$) truncated at a radius r_{max}, r_{max} can be larger than r_{arc} and $\Sigma_o > \Sigma_c$. The source size is then equal to the width of the arc independent of other lens parameters. Because of this, the above difficulties can be avoided but r_{max} still must be less than the extent of the X-ray emitting gas ($>$ Mpc) for M/L_B not to exceed $1000 M_\odot / L_\odot$. The severity of the difficulties can also be reduced by a nonspherical distribution of dark matter [10,11] but with added free parameters for the lens model.

B) Source Structure

The following conclusion can be drawn about the source structure.

1. The source behind Abell 370 must be a fairly elongated galaxy to produce the observed deviation of the arc from circularity [1,8] and the evident variation of the width from West to East [1,11]. This conclusion is based on models used for construction of images [8] and is evident from attempts to reconstruct the source from the detailed distribution of the image surface brightness [12].

2. The image construction of Cl2244 indicates the presence of a nearly spherical source with fairly high surface brightness (uncomfortably high for a galaxy if the higher redshift, $z_s = 2.2$, turns out to be correct). The well defined surface fluctuations [1,13] can be very useful for source reconstruction which is being carried out now. Whether the fluctuations are due to the merging of two images for a triple source (a spiral galaxy with well defined core and two arms [9]) or due to the merging of three images of a double source or due to mini lensing by small invisible galaxies or clumps in the dark matter distribution [8] remains to be seen.

III. PROBABILITY AND STATISTICS

Let $p(m)dm$ be the probability of magnification in the range m to $m + dm$ and $P(m) = \int_m^\infty p(m')dm'$ the cumulative probability of magnification greater than m. Then the rate of occurence of arcs or images with a certain magnification will be $R(m) = R_o p(m)$, where R_o determines the absolute value of the rate and $p(m)$ describes the relative probabilities. Both of these depend in a complicated way on the selection of sources and lenses, on the structure of lenses, and on various cosmological parameters.

As mentioned above there are two difficulties with the interpretation of the rate of occurence of the arcs. The first is that the expected absolute value of the rate for arcs as highly magnified as the ones in Abell 370 or Cl2244 is lower than observed (calculated value of R_o is too small by a factor of 10 or larger). In a well defined sample (i.e. with known surface brightness and magnitude limits for sources and known limits on clusters

of galaxies), the absolute probability depends on essentially two factors. The first is the fraction of sky covered with sources within the chosen limits and the second is the surface mass density distribution of clusters divided by the critical density Σ_c (i.e. the distribution of central or mean surface densities $\sigma_o = \Sigma_o/\Sigma_c$ or $\bar{\sigma} = \overline{\Sigma}/\Sigma_c$). These factors are not known very precisely and a detailed analysis of them is beyond the scope of this report and will be treated in a future work. (Some elementary discussion of these can be found in references 7, 14, and 15). We assume here that the discrepancy of a factor of ten or higher in the value of R_o can be resolved by better understanding of these factors and the properties of the more complex lenses needed for modeling of the arcs. Some of the discrepancy could also be due to statistical fluctuations associated with small size samples or in other words there was some luck involved in the initial discovery of the two well defined and highly magnified arcs [1].

Be that as it may, we concentrate here on the second difficulty, namely the relative probabilities or the shape of the probability function $p(m)$. The problem here is that for simple lenses at $m \gg 1$ we expect $p(m) \propto m^{-3}$ or $P(m) \propto m^{-2}$, so that $P(5) : P(10) : P(20) = 16 : 4 : 1$, which is not observed. In what follows we derive the probability distributions for various lens models.

A) Spherically symmetric lenses

Following the notation in reference 8 we write the lens equation in terms of the angles ϕ and θ in the source and lens plane as

$$\vec{\phi} = \vec{\theta} - \vec{\tilde{\alpha}}(\vec{\theta}), \tag{1}$$

where $\vec{\tilde{\alpha}}$ is the deflection angle $\vec{\alpha}$ multiplied by D_{sl}/D_s and is given by

$$\vec{\tilde{\alpha}} = \frac{1}{\pi} \int \sigma(\vec{\theta'}) \frac{\vec{\theta} - \vec{\theta'}}{|\vec{\theta} - \vec{\theta'}|^2} d\theta'_x d\theta'_y. \tag{2}$$

For a spherically symmetric surface density distribution $\cdot\sigma(\theta) = \sigma_o f(\theta/\theta_o)$, $\tilde{\alpha}$ has only a radial component with magnitude

$$\tilde{\alpha}(\theta) = \frac{\sigma_o}{\pi}\frac{1}{\theta}\int_0^\theta 2\pi f(\theta'/\theta)\theta' \, d\theta' = \frac{M(D_l\theta)}{M_c\theta}, \quad M_c = \pi\Sigma_c D_l^2. \tag{3}$$

The magnification m of an infinitesimal source is given by the inverse of the determinant of the Jacobian $\partial\vec{\theta}/\partial\vec{\phi}$:

$$m^{-1} = m_t^{-1}m_r^{-1} = \left|\left(1 - \frac{\tilde{\alpha}(\theta)}{\theta}\right)\left(1 - \frac{d\tilde{\alpha}}{d\theta}\right)\right|. \tag{4}$$

Here the first term on the right describes the inverse of the tangential magnification m_t and the second the inverse of the radial magnification m_r.

1. *Point mass lens (pm)*: For a lens with all its mass M concentrated within a small region about the center of coordinates $\tilde{\alpha} = M/(M_c\theta)$ and there are always two images with magnification $m^{-1} = \pm(1 - \theta_e^4/\theta^4)$, where θ_e is the radius of the Einstein ring and the $\pm$ signs correspond to the positive and negative parity images. Using equation (1) to eliminate θ and solving for ϕ, we find $\pi\phi^2 = \pi(M/M_c)[(1 \pm m^{-1})^{-1/4} + (1 \pm m^{-1})^{1/4}]^2$ which for $m > 1$ is proportional to the cumulative probability $P(m)$ for each parity image because all sources within this area will have magnification greater than m. For $m < 1$, the solution with the minus sign is not possible (so that the positive parity image cannot have $m < 1$) and our expression with the plus sign then corresponds to the cumulative probability for having a magnification between 0 and m. Differentiation of $P(m)$ gives

$$p_{pm}(m) = \frac{\pi M}{2M_c}m^{-3/2}((m-1)^{-3/2}\Theta(m-1) + (m+1)^{-3/2}). \tag{5}$$

Here $\Theta(x) = 1$ for $x > 0$ and is zero for $x < 0$. The first term corresponds to the contribution to the probability from the positive parity image while the second is the probability derived for the negative parity image.

2. *Isothermal Gas Sphere (igs)*: The isothermal gas sphere strictly speaking has infinite surface mass density at the center. In this case the bending angle is $\tilde{\alpha} = \alpha_o$ with

$\alpha_o = v^2/(2\pi G\Sigma_c D_l)$, where $v = (2GM/r)^{1/2}$ is the constant velocity dispersion. For source postions $\phi < \alpha_o$ there are two images but for $\phi > \alpha_o$ the negative parity image does not exist. Following the above procedure we obtain the probability distribution

$$p_{igs}(m) = 2\pi\alpha_o^2((m-1)^{-3}\Theta(m-1)+(m+1)^{-3}). \tag{6}$$

As shown to be the case for more general lenses [14,16], both of the above probability distributions vary as m^{-3} for large magnifications .

3. Infinite uniform sheet (ius): If the surface density is nearly constant to distances larger than under consideration, one can use the approximation $\theta_o \to \infty$ and $f(\theta/\theta_o) \to 1$ so that from (3) and (4) we have $\tilde{\alpha} = \sigma_o\theta$ and $m^{-1} = (1 - \sigma_o)^2$. Here for given σ_o there is only one image with magnification $m_o = (1 - \sigma_o)^{-2}$, so that $p_{ius} \propto \delta(m - m_o)$. In this case the actual probability distribution will be determined by the distribution of σ_o (i.e. both Σ_o and Σ_c)[16,17].

4. More general lenses : For more realistic lenses with finite values of Σ_o and θ_o the deflection angle is expected to initially increase (like for a uniform sheet), to reach a maximum (if the maximum is broad, the bending angle behaves like in an *igs* model) and then to decrease while approaching the point mass case at large distances ($\theta \gg \theta_o$). The Gaussian distribution and the distribution termed simple (S) in reference 8 satisfy this prescription. In general, if we write the deflection angle as

$$\tilde{\alpha} = \sigma_o\frac{\theta}{1 + (\theta/\theta_o)^n}, \tag{7}$$

then $n = 0, 1,$ and 2 correspond to the *ius*, *igs* (non-singular), and S models, respectively. If we let $\theta_o \to \infty$, we obtain the *ius* and if $\theta_o \to 0$, we have for $\theta > \theta_o$ a point mass model. The deflection angles (7) for $n = 1$ or 2 do not correspond to a specific physical model (although their general shape is what one would expect from physically reasonable mass distributions) but behave smoothly (no singularities at the origin) and the magnifications can be derived analytically. Indeed, the magnification equation (4)

reduces to a cubic in $z = r^n + r_o^n$ with coefficients that depend on α_o, θ_o, and the magnification.

We will discuss these solutions in general since the detailed algebraic form varies with n while the physical interpretation is essentially the same for all $n \leq 2$ or even more complicated lenses with the same general behaviour described above. We demonstrate that considerable deviation of the probability distribution $p(m)$ from the simple monotonic m^{-3} dependence (such as in equations 5 and 6) can be obtained. As is well known and can be verified from equation 4, for a distributed lens such as that described by equation 7 there are two critical lines. One of these, corresponding to infinite tangential magnification, is at the Einstein radius $\theta_e = \theta_o(\sigma_o - 1)^{1/n}$, and the other corresponding to infinite radial magnification is at a radius θ_r. The variation with θ of the total magnification (4) and the ratio of the tangential to radial magnification $\xi = m_t/m_r$ are plotted in Figure 1 for $n = 2$, $\sigma_o = 1.91$, and $\theta_o = 1.2$.

For source positions $\phi > \phi_1 = \theta_1 - \tilde{\alpha}(\theta_1)$ there is only one image at $\theta > \theta_1$, with θ_1 defined by $\theta_1 - \tilde{\alpha}(\theta_1) = \tilde{\alpha}(\theta_r) - \theta_r$. This image has magnification $m < m_{min}$ and is primarily tangentially magnified so that $\xi(\theta)$ satifies $\xi(\theta) > 1$. The probability $p(m)$ for this image is obtained by differentiating the area $\pi\phi^2$ with respect to m. For $0 < \phi < \phi_1$, there are three images, one of which is described by the continuation of the right branch in Figure 1 from θ_1 to θ_e, while the other two start for $\phi = \phi_1$ at θ_r (with infinite radial magnification) and follow for smaller ϕ the two branches to $\theta = 0$ and $\theta = \theta_e$ as $\phi \to 0$ (as indicated by the arrows). We can evaluate the probability distribution for these images from the respective areas in the source plane. We note that, for $n = 2$, the minimum magnification occurs at $\theta_{min}/\theta_o = \sqrt{2(\sigma_o - 1)/(\sigma_o + 2)}$ and $m_{min}^{-1} = (\sigma_o - 1)(\sigma_o^2 + 7\sigma_o - 8)/(27\sigma_o)$. In general the minimum magnification m_{min} is equal to $m(\theta_1)$ since both seperate the region at which there is one solution from the region at which there are three solutions. For given m such that $m > m_{min}$, there are four solutions in the lens plane, so that four source positions give us contributions to the probability distribution at these values of m. For $m < m_{min}$, the middle branch of
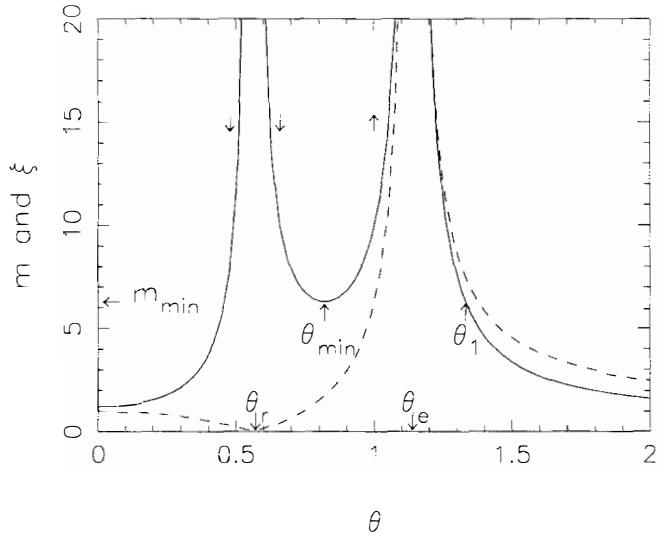
Figure 1: Variation of total magnification $m = m_r m_t$ (solid line) and the magnification ratio $\xi = m_t/m_r$ (dashed line) with position θ of the images in the lens plane for the model of equation (7) with $n = 2$ (see text Section III.A.4).

Figure 1 does not contribute and there are only two contributions to the probability. Thus there is a discontinous increase in the probabilty distribution at $m > m_{min}$. This discontinuity increases the cumulative probability $P(m_{min})$ by approximately a factor of two over that which would be expected from a simple extrapolation of the m^{-3} approximation.

The discontinuity in the probability is due to the presence of a distinct radial critical line near which the images are stretched in the radial direction. However, we are not interested in the probability of *all* highly magnified images, but only those with large tangential magnifications. Therefore, to analyze the probability of producing tangentially elongated images we must also consider the variation of magnification ratios ξ, shown in Figure 1.

It is evident that tangentially magnified images (with $\xi > 3$ so that the arc or arclet will be distinguishable from a lenticular unmagnified galaxy) are associated with image positions $\theta_{min} < \theta < \theta_1$ and have magnifications greater than m_{min}. This is precisely the region of enhanced probabilities. We therefore conclude that even with the simplest realistic (not singular like those in items 1, 2, and 3 above) surface density distributions the probability distribution could deviate from the m^{-3} law significantly at high magnifications. It is clear from Figure 1 that by simple adjustment of the parameter θ_o and σ_o this enhancement can be optimized to produce its largest effect near images of size $\approx 20''$ ($m \approx 20$). To show this, we plot in Figure 2 the ratio of probabilities $p(20, \xi > 10)/p(5)$ divided by the same quantity for a point mass as a function of σ_o. (The ratio is independent of θ_o). Note that for $\sigma_o \approx 1.7$, we get a higher ratio than for the point mass. At even higher σ_o , m_{min} shifts to a lower value and the probability distribution is enhanced at lower values of m so that the curve shows a minimum at the value of σ_o which corresponds to a high probability of $m = 5$. It would be important to ascertain if astrophysical data favors values of σ_o and θ_o which correspond to a large value of the above ratio. There is some evidence that the lower
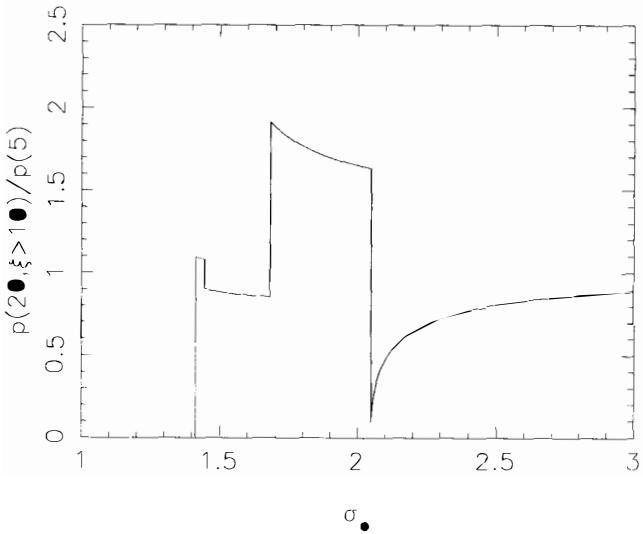
Figure 2 The ratio $p(20, \xi > 10)/p(5)$ is plotted versus σ_o for the lens model described by equation (7) with $n = 2$. The ratio unity corresponds to the ratio expected for a point mass lens (see text Section III.A.4).

values of σ_o are more likely. If so, the discrepancy between the expected and observed relative probabilities could be decreased significantly.

B) Extended sources

The above probabilities are for infinitesimal sources and not accurate at very high magnifications. In general, the source will be extended either by being a resolvable source (when not lensed) or due to seeing. The maximum magnification of a tangentially elongated image will be at $\phi = 0$. At that position, the magnification, given by the ratio of area of the image to that of the source, is $m_{max} \approx \pi \theta_e \delta \theta / \pi \delta \phi^2$, where $\delta \phi = \delta \theta (1 - d\tilde{\alpha}/d\theta|_{\theta_e})$ is the radius of the source (assumed to be a circular disk) as would be observed without the presence of the lens and $\delta \theta = 2\pi \theta_e / \xi$ is the observed width of the image. If $\delta \theta$ and $\delta \phi$ are approximately the size of the seeing disk θ_s, $m_{max} \approx \theta_e / \theta_s$. For reasonable values of the parameters σ_o and θ_o, $\theta_e \approx 30''$, so that the probability curves are truncated at $m_{max} \approx 30$ with the cumalitive probability $P(m_{max})$ added to the value of probabilities just below m_{max}, consequently enhancing the probability of observing $m \approx m_{max}$. This also will alleviate the above discrepancy.

We note in passing that the probability distributions break down at the other limits ($m \to 1$ or $m \to 0$) as well. Otherwise not only the differential but also the cumulative probabilities would diverge (see equations 5 and 6). This is of course not physical and arises because of the assumptions of infinite thin flat planes for the source and the lens. Small values of m correspond to large values of source and image positions ϕ and θ while the small angle approximations used in deriving the lens equation (1) are valid for ϕ and $\theta \ll 1$. Thus we can ignore the divergences at $m \to 1$ or $m \to 0$.

C) Non-spherical Lenses

We have so far discussed spherically symmetric lenses. None of the models used to produce the giant luminous arcs were spherically symmetric since a spherically symmetric lens will produce two arcs. The deviation of the absolute probability distribution from the simple models due to asphericity in the lens is beyond the scope of this report. Here we present only results from a simple numerical simulation which we have carried
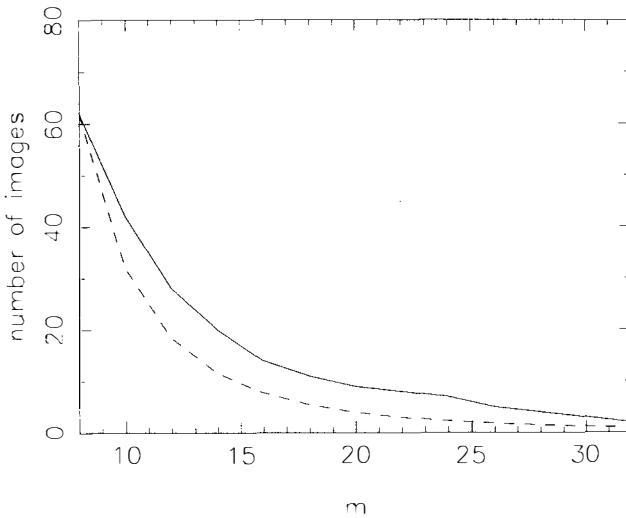
Figure 3: Differential probability distribution for a point mass (dashed line) and a lens model with Gaussian surface density distribution (solid line) which produces a good image of the arc in Abell 370 (from reference 8). The two curves are normalized at $m = 8$ (see text Section III.C).

out using one of the lens models which is able to produce the observed arc in A370. We move a source behind the lens and compare the number of images with different lengths (or magnifications). In Figure 3 we show the result of these numerical simulations (solid line) for a Gaussian mass distribution with characteristic size somewhat smaller than the angular extent of the arc ($\theta_o < \theta_{arc}$). The dashed line shows the expected distribution for the $p(m) \propto m^{-3}$ approximation. As evident, the probability for the complex lens is higher by a factor of approximately 2 and the relative probability between images of $m \approx 10$ to those with $m \approx 20$ is also increased by a factor of 2.

In summary we see that various factors in realistic lenses can enhance the probability of occurence of highly elongated images over that expected for simple lenses. Whether these will be the correct explanation will require further theoretical work and a larger sample of observed arcs and arclets.

References

[1] Lynds, R., and Petrosian, V. 1989, *Ap. J.*, **336**, 1.

[2] Soucail, G., Mellier, Y., Fort, B., Mathez, G., and Cailloux, M. 1988, *Astr. Ap.*, **191**, L19.

[3] Soucail, G., Mellier, Y., Fort, B., Mathez, G., and Cailloux, M. 1989. In *The Toulouse Workshop on Gravitational Lensing*, ed. Y. Mellier, G. Soucail, and B. Fort. Springer Verlag.

[4] Pello, R., Le Borgne, J.F., Mathez, G., Mellier, Y., Sanahuja, B., and Soucail, G. 1989. In *The Toulouse Workshop on Gravitational Lensing*, ed. Y. Mellier, G. Soucail, and B. Fort. Springer Verlag.

[5] *The Toulouse Workshop on Gravitational Lensing*, ed. Y. Mellier, G. Soucail, and B. Fort 1989, Springer Verlag.

[6] *Gravitational Lenses, Lecture Notes in Physics, vol. 330*, ed. J.M. Moran, J.N. Hewitt, and K.Y. Lo, Berlin: Springer-Verlag.

[7] Nemiroff, R.J., and Dekel, A. 1989, *Ap. J.*, **344**, 51.

[8] Bergmann, A.G., Petrosian, V., and Lynds, R. 1990, *Ap. J.*, **350**, 23.

[9] Hammer, F., and Rigaut, F. 1989, *Astr. Ap.*, **226**, 45.

[10] Grossman, S., and Narayan, R. 1989, *Ap. J.*, **344**, 637.

[11] Mellier, Y., Soucail, G., Fort, B., Le Borgne, J.F., and Pello, R., 1989. In *The Toulouse Workshop on Gravitational Lensing*, ed. Y. Mellier, G. Soucail, and B. Fort. Springer Verlag.

[12] Petrosian, V., Bergmann, A., and Lynds, R. 1989. In *The Toulouse Workshop on Gravitational Lensing*, ed. Y. Mellier, G. Soucail, and B. Fort. Springer Verlag.

[13] Hammer, F., Le Févre, O., Jones, J., Rigaut, F., Soucail, G. 1989, *Astr. Ap.*, **208**, L7.

[14] Blandford, R.D., and Narayan, R. 1986, *Ap. J.*, **310**, 568.

[15] Grossman, S., and Narayan, R. 1988, *Ap. J. (Letters)*, **324**, L37.

[16] Kovner, I. 1987 , *Ap. J.*, **321**, 686.

[17] Turner, E.L, Ostriker, J.R., and Gott, J.R. 1984, *Ap. J.*, **284**, 1.

LOWER MASS-LIMITS ON DARK MATTER FERMIONS AND BOSONS.

Jes Madsen
Institute of Astronomy
University of Aarhus
DK-8000 Aarhus C, Denmark.

ABSTRACT

The derivation of lower mass-bounds for primordial neutrinos constituting dark matter in galaxies is discussed on the basis of the Tremaine-Gunn phase-space constraint and extensions thereof. A number of problems related to the unknown nature of the coarse-grained phase-space distribution are mentioned. It is also shown how similar phase-space constraints can be applied to primordial bosons.

I. THE TREMAINE-GUNN CONSTRAINT ON m_ν.

The idea that massive neutrinos could constitute the dark matter around galax-
ies seems to date back to a paper by Markov[1] from 1964. Cowsik & McClel-
land[2] and Marx & Szalay[3] demonstrated, that neutrinos in a degenerate confi-
guration with particle mass of the order of 10 eV fitted well with observed
masses and radii of galaxy clusters, and several groups refined and extended
these calculations to describe galaxy halos as well, notably a series of papers
by Ruffini and coworkers calculating properties of degenerate spheres with zero
and non-zero temperature (e.g. ref.4).

A basic feature of most of these early investigations (and a number of
later ones as well) was the fact that a degenerate neutrino sphere at tempera-
ture T=0 has a mass-radius relation

$$M = 1.1 \times 10^{11} M_\odot R_{10}^{-3} m_{\nu 30}^{-8} g^{-2} \tag{1}$$

where $R_{10} \equiv R/10\text{kpc}$, and $m_{\nu 30} \equiv m_\nu/30\text{eV}$. Equation (1) assumes one flavour of
neutrinos and antineutrinos, and $g \approx 1$ is the effective number of helicity states.
A neutrino mass $m_{\nu 30} \approx 1$, close to the experimental limit for electron antineutri-
nos (and for a time even a "measured" value), gives a nice fit to galaxy radii
and masses.

However, Equation (1) is not directly useful as neutrino halos cannot be
degenerate. The observational evidence for this is the fact that the density
structure of a degenerate sphere gives rise to a rotation curve which is far
from flat, as opposed to observed rotation curves. The theoretical reason is the
fact, that the primordial neutrino distribution function is far from degeneracy
when neutrinos decouple from equilibrium with the photons, at a temperature of
order 1 MeV. For dissipationless evolution it is not possible to compress neutri-
nos in phase-space (Liouville's theorem). The coarse-grained distribution function
relevant for dark halos is created by mixing of the fine-grained distribution
function, and one can quantify the distribution of coarse-grained phase-space
density by using the procedure described by Tremaine, Hénon & Lynden-Bell[5]
(see also ref.6). In particular, the *maximum* coarse-grained phase-space density
cannot exceed the *maximum* fine-grained phase-space density of the primordial
distribution, which is 0.5 in units of g/h^3. This was realized by Tremaine &
Gunn[7] and used for deriving a lower bound on the neutrino mass in galaxies
assuming the coarse-grained neutrino distribution to be an isothermal sphere
with core-radius R_c and one-dimensional velocity dispersion σ. On the premise
that the central (maximal) phase-space density of the isothermal sphere should
not exceed g/h^3, where a factor of 2 has been included to account for one fla-
vour of particles and antiparticles, they showed that

$$m_\nu > (38\text{eV})\sigma_{100}^{-1/4}R_{c10}^{-1/2}g^{-1/4}, \tag{2}$$

where $\sigma_{100} \equiv \sigma/100\text{kms}^{-1}$, and $R_{c10} \equiv R_c/10\text{kpc}$.

The Tremaine-Gunn limit has been applied to a number of galaxies by a number of authors. "Typical" neutrino mass limits derived for spiral galaxies are in the range from 5-40 eV, whereas the presence of dark matter in dwarf spheroidals (still a somewhat controversial issue on the observational side) would correspond to limits in the range from 100-300 eV. For comparison the upper cosmological bound for a single stable flavour[8] is $m_\nu \leq 91.5\text{eV}\Omega h_0^2 g^{-1}$, and the experimental limit on the mass of the electron antineutrino is[9] 18 eV. Different authors have drawn different conclusions on the viability of neutrino dark matter from these numbers.

One should be aware, that the application of Equation (2) is subject to a number of problems.

1. It is very difficult to subtract the luminous contribution from most rotation curves, thereby isolating the dark matter contribution. Only a quite small number of galaxies are modeled in such detail, that it makes sense to extract well-defined parameters for the dark halo. Better observations and detailed modeling have shown, that many of the galaxies for which neutrino mass-limits have been published (including some used by the present author) are unsuited for that purpose (e.g. references 10 and 11).

2. The dark matter is (by definition) not directly observable. The actual coarse-grained phase-space distribution cannot be seen, and there is no guarantee, that the assumption of an isothermal sphere, which is a basis of Equation (2), is correct.[12-15] In fact it is not even known whether the dark matter has an isotropic velocity distribution. As discussed below, deviations from isotropy may significantly change mass-limits, thereby weakening the Tremaine-Gunn constraint.

3. If an isothermal sphere is assumed for the dark matter, there are still problems related to the determination of the core radius and velocity dispersion. Using measured values for the luminous component is not necessarily correct. In fact there are good reasons for believing, that the dark matter is more extended than the luminous matter, so use of the luminous core radius makes the neutrino mass limits too strong. Similar problems relate to the velocity dispersion. For instance the neutrino mass determined from dwarf spheroidals may be reduced to acceptable values if the dark halos are much more extended than the stellar component.[12),14-16]

4. Finally one should incorporate knowledge about the complete fine-grained distribution function, and the corresponding bounds on the coarse-grained

distribution, instead of only comparing maximum phase-space densities. An attempt in this direction was made by Madsen & Epstein[12], who introduced the concept of a maximally compact sphere of neutrinos consistent with the neutrino distribution function at decoupling and the spherical Jeans-equation (the construction is not a steady-state solution, but nevertheless gives useful instantaneous limits).

The effects of anisotropies in the velocity distribution can be illustrated by means of a simple example taken from ref.17. Consider dark matter with a distribution function of the form

$$f(\varepsilon, L) = constant \times L^{-2\alpha} \varepsilon^{n-3/2}, \qquad (3)$$

where ε and L are the relative energy and the angular momentum and n and α are constants. These are the generalized polytropes first introduced by Hénon.[18] Spherical systems with $\alpha > 0$ have diverging densities in the center. Systems with $\alpha = 0$ are the usual isotropic polytropes, and for these $n=3/2$ is the degenerate sphere. For $\alpha < 0$ transverse velocity dispersion dominates, and the configurations are shell-like with zero mass-density in the center. Fixing the maximum phase-space density of the particles (to equal the maximum of the primordial fine-grained density, which cannot be exceeded during dissipationless evolution), one gets the interesting result, that all solutions with $n \gtrsim 3/2$ and $\alpha < 0$ have *larger* masses for fixed radius and particle mass than the degenerate sphere. Most of this mass is situated at large radii. It is not claimed that the dark matter is described by generalized polytropes, but the example serves to illustrate, that anisotropic systems can be more massive than isotropic ones, thus relaxing constraints on particle masses. Furthermore most of the mass could be situated at large radii without "spoiling" the inner parts of the rotation curves, that are more easily explained by luminous material.

Anisotropies in the velocity distribution can be dealt with in many different ways. For instance Madsen & Epstein[13] used an equation of state approach to model the most compact distribution of neutrinos that could be constructed for different degrees of anisotropy in the *outer* regions of halos. The equation of state approach does not give time-independent solutions to the Jeans-equations, but gives an instantaneous conservative limit on the neutrino mass. Recently Ralston & Smith[19] have approached the anisotropy-issue from a somewhat different point of view.

II. BOSON MASS LIMITS.

It is worth stressing, that it is *not* the Pauli principle, but rather the non-increase of the maximal occupation number during dissipationless evolution that lies behind the Tremaine-Gunn constraint. At first sight the difference may only seem to be a factor of $2^{1/4}$, but the logical difference is important. Thus a gas of bosons with zero chemical potential obeys no Pauli principle, and there is no maximum occupation number. The latter point led Tremaine & Gunn[7] to conclude that consequently similar constraints did not apply to bosons. However, as nicely illustrated by Tremaine, Hénon & Lynden-Bell,[5] there are restrictions on the evolution of the complete phase-space distribution, not only the maximum occupation number, and this results in boson mass-limits very similar to those for neutrinos, as was recently shown by Madsen[6] [Turner[20] briefly touched upon related ideas in connection with thermal axions]. The crucial difference is, that the boson-limit is statistical, i.e. it must be applied to a sample of galaxies.

Primordial bosons decoupling from thermal equilibrium in the relativistic regime have an occupation number distribution of the form $f_{BR}(p) \approx 1/(\exp(pc/kT)-1)$, where p is the momentum. (This assumes zero chemical potential; more on non-zero chemical potential and non-relativistic decoupling below and in ref.6). f_{BR} diverges for small momenta, but the important point is, that only a very small fraction of the particles have large occupation numbers. Less than 10% of bosons decoupling in the relativistic regime have fine-grained phase-space density exceeding $0.8g/h^3$, and less than 1% exceeds $4g/h^3$. Dissipationless mixing allows an increase in the fraction of particles that may end up in regions with (coarse-grained) phase-space densities exceeding a given number. The detailed fine- and coarse-grained distributions are given in ref.6, calculated on the basis of the mixing theorem recipe in ref.5. In the case of relativistic decoupling, the coarse-grained phase-space densities corresponding to 10% and 1% of the particles are $2g/h^3$ and $6g/h^3$, representing only a mild increase relative to the fine-grained distribution. The mean value of the occupation number is 2.4 for the densest 10% fraction. Notice that these numbers are not very much higher than the maximum value $0.5g/h^3$ for the fermion distribution.

Parametrizing the "typical" boson fine-grained phase-space density as $\tau g/h^3$, and comparing this to the maximal coarse-grained phase-space density for an isothermal sphere, one obtains a relation very similar to Equation (2), putting a lower bound on the boson mass of

$$m_B > (38\text{eV})\sigma_{100}^{-1/4}R_{C10}^{-1/2}g^{-1/4}\tau^{-1/4}. \tag{4}$$

Since only the fourth root of τ enters in the mass-limit, the limits derived for bosons are only slightly smaller than the corresponding limits for fermions. This holds true also for non-relativistic decoupling, even though the relevant values of τ in that case may be several hundreds.[6]

The boson mass-limits derived from Equation (4) are statistical in the sense, that they apply to a sample of galaxies (if one accepts the assumption of an isothermal sphere for the coarse-grained distribution; of course all the problems noted above in the neutrino-case apply for bosons as well). For fermions the existence of a maximum occupation number means, that Equation (2) can be applied to individual galaxies, but for bosons some halos might be constructed from extremely high-occupation-number bosons. However, as 10% or more of the dark matter must be incorporated in galaxy halos, on the average the halos must consist of low-density bosons. Selection effects can thus be avoided by using large samples.

Zero chemical potential is an extreme situation. If bosons are their own antiparticles and annihilate into photons, they must have $\mu = -m_B c^2$, and in general the chemical potential of a boson able to annihilate with its antiparticle into photons must have $-2m_B c^2 \le \mu \le 0$ (restmass-terms are consistently subtracted in the definition of energy and chemical potential). Thus in most cases $\mu \ne 0$, and so the occupation number has a maximum of $1/(\exp(|\mu|/kT)-1)$, calculated at the decoupling temperature. This number can be set equal to τ to get a firm, "non-statistical" mass-limit, but "statistical" limits based on the complete distribution function will still be much stricter for relativistic decoupling. (This is because a non-zero chemical potential is limited by $-2m_B c^2 \le \mu \le 0$ so that μ/kT is small at the relativistic decoupling temperature. For the same reason the "statistical" limits obtained from the full $\mu \ne 0$ distributions are almost identical to those for $\mu = 0$). For non-relativistic decoupling the non-statistical mass-limits are of more use, but again one should notice that if bosons are distinct from their antiparticles, μ/kT can be very small for either boson or antiboson.

At present only the axion seems to be a respectable bosonic candidate for the dark matter, and the limit above does not apply to the standard coherent axion or axions from decaying strings, both of which are non-thermal in origin. The limit applies only to thermal axions and other bosons once in thermal equilibrium.

III. CONCLUSION.

Constraints on the coarse-grained phase-space distribution that can be reached by dissipationless processes from a fine-grained distribution at decoupling leads not only to the well-known Tremaine-Gunn limit on neutrino- (fermion-) masses, but also to rather similar limits for dark matter bosons. The boson mass-limits are statistical in nature due to lack of a maximum occupation number (for zero chemical potential); i.e. they should be applied to a sample of galaxies rather than to individual galaxies. "Non-statistical" limits can be obtained for non-zero chemical potential.

The limits given as Equations (2) and (4) above are however subject to the important assumption, that the dark matter distribution in halos takes the form of an isothermal sphere. As discussed in Section I, mass-limits are generally reduced when deviations from this shape for the distribution function are taken into consideration. But for order-of-magnitude estimates the equations give very useful bounds on the masses of dark matter particles, be it fermions *or* bosons, assuming only that these particles evolved dissipationlessly from a thermal equilibrium distribution at decoupling.

REFERENCES.

1)   M.A.Markov, Phys.Lett.**10**, 122 (1964).

2)   R.Cowsik and J.McClelland, Astrophys.J.**180**, 7 (1973).

3)   A.S.Szalay and G.Marx, Astr.Ap.**49**, 437 (1976).

4)   R.Ruffini, Lett.Nuov.Cim.**29**, 161 (1980).

5)   S.Tremaine, M.Hénon, and D.Lynden-Bell, Mon.Not.Roy.Astron.Soc.**219**, 285 (1986).

6)   J.Madsen, Phys.Rev.Lett.**64** (1990) (in press).

7)   S.Tremaine and J.E.Gunn, Phys.Rev.Lett.**42**, 407 (1979).

8)   R.Cowsik and J.McClelland, Phys.Rev.Lett.**29**, 669 (1972).

9)   M.Fritschi *et al.*, Phys.Lett.**B173**, 485 (1986).

10)   S.M.Kent, Astron.J.**93**, 816 (1987).

11)   G.Lake, Astron.J.**98**, 1253 (1989).

12)   J.Madsen and R.I.Epstein, Astrophys.J.**282**, 11 (1984).

13)   J.Madsen and R.I.Epstein, Phys.Rev.Lett.**54**, 2720 (1985).

14)   R.Cowsik and P.Ghosh, Astrophys.J.**317**, 26 (1987).

15)   J.P.Ralston, Phys.Rev.Lett.**63**, 1038 (1989).

16)   J.Kormendy, Proceedings IAU Symposium **117**, 139 (1987).

17)   J.Madsen, Astrophys.J. (submitted) (1990).

18)   M.Hénon, Astr.Ap.**24**, 229 (1973).

19)   J.P.Ralston and L.L.Smith, Kansas preprint (1990).

20)   M.S.Turner, Phys.Rev.Lett.**59**, 2489 (1987).

COSMIC RINGS
AS A CHUMP DARK MATTER CANDIDATE?

Brandon CARTER
Département d'Astrophysique Relativiste et de Cosmologie
C.N.R.S. - Observatoire de Paris
92195 MEUDON, France.

Abstract.

For spontaneous symmetry breaking at mass scales comparable with the 10^{16} G.e.v. range predicted for grand unification, the suggestion that a resulting network of cosmic strings (i.e. vortex defects of the vacuum) might have been of of the superconducting kind would appear to be excluded by the consideration that they would in that case have given rise to a cosmological mass excess due to the formation of "cosmic rings", i.e. circular rotating equilibrium states in which the energy of a string loop attains a minimum for given values of the conserved charge number C and the topological winding number N whose product determines the angular momentum quantum number $J = CN$ of the ring. A tentative preliminary estimate indicates that the critical symmetry breaking mass scale below which the resulting cosmic ring distribution would be cosmologically unimportant is very roughly (within a few powers of ten) of the order of 10^4 G.e.v. If the symmetry breaking actually occurs at a mass scale comparable with this critical value the resulting cosmic rings could provide an important "chump" (i.e. charged ultra massive particle) contribution to the dark matter of the universe. Such rings would typically have charge numbers of the same order as those of ordinary chemical elements, $\approx 10^2$, and their typical masses would exceed the relevant symmetry breaking mass scale by a factor of the same order. The resulting ring mass estimate of $\approx 10^6$ G.e.v. is remarkably comparable with previously claimed lower limits for the masses of chump dark matter candidates.

The purpose of this communication is to draw attention to a kind of "chump" that may conceivably fulfil the conditions for being a cosmologically important dark matter candidate, but that has so far received less serious consideration than it would seem to deserve, as the possibility of its existence does not appear to have been conceived before the recent work of David and Shellard[1]. The chumps in question are the *cosmic rings* that, it now seems, would occur genericly[2][3] as ground state equilibrium configurations of superconducting string loops of a broad category[4][5] including the various kinds proposed by Witten[6] and many subsequent workers as vortex defects of the vacuum in field theories with spontaneous symmetry breaking. No such equilibrium states exist in the more thoroughly studied case of non conducting string loops, which in the long run must decay to nothing by gravitational radiation. However although electromagnetic radiation can initially bring about even faster energy loss in the case of conduction string loops, the completion of their decay will generically be prevented by the existence of non-vanishing conserved charge and topological winding numbers, so that durable relics will be left over in the form of rings.

Davis and Shellard remarked[1] that for symmetry breaking at the order of the "grand unification" mass scale $\approx 10^{16}$ G.e.v. the ensuing relics would give rise to a cosmological mass excess problem analogous to the familiar monopole mass excess problem. The present work contains a first tentative estimate of the maximum symmetry breaking mass scale below which this problem would not arise, so that the resulting rings would not necessarily have to be "inflated away" but might conceivably be present as a constituent of the universe now. It turns out rather remarkably (as a coincidence of conceivably "anthropic" significance?) that the consequent estimate of the maximum allowable value for the typical ring mass is of the same rough order (to within the uncertainties which at this tentative stage are of several powers of ten) as the lower mass limits that have been estimated as being presently conceivable for "chump" dark matter candidates, being something round about 10^6 G.e.v.

The possibility of "chumps" as an alternative to the more commonly discussed possibility of "wimps", meaning weakly interacting massive particles, for the role of the cosmologically dominant intergalactic dark matter contribution, has been previousl discussed[7][8] without specific reference to string loops. We use the term *chumps* as an abbreviation for charged ultra-massive particles. The qualification *ultra-massive* is essential if the weak interaction restriction is to be dispensed with, and is to be interpreted in this context as meaning not too far below the T.e.v. regime, depending on the relevant cross section. In order to be cosmologically important, particles with more moderate masses, in the G.e.v. regime or below, must have such a high number density n that the weak interaction qualification is necessary, to ensure that their scattering cross sections are small compared with the geometric cross section of an atomic nucleus, which is necessary for their collision timescale to be long compared with the age of the universe[7]. In the case of electrically charged particles, even though they would be expected to be shielded, the effective cross section would be much larger, at least in the positively charged case for which the sheilding would presumably be provided by electrons in a configuration of the familiar atomic kind characterised by the ordinary Bohr radius scale, so that the corresponding minimum mass would be proportionally increased. It is however to be born in mind that the charges and currents of conceivable kinds of superconducting string loop do not necessarily have to be of the electrically coupled kind but could well be of a neutral variety (such as that made familiar by the standard Glashow Weinberg Salam theory) in which case the less severe limit, of the order of tens of G.e.v. rather than thousands of T.e.v., would apply.

When the cosmological temperature Θ drops to the critical value

$$\Theta_x \approx m_x \qquad (1)$$

where m_x is the relevant (provisionally unspecified) symmetry breaking mass scale (we use Plank units with $c = G = h/2\pi = 1$) one expects[9][10] the formation of a network of string defects with initial collective mass density

$$\rho_x \approx m_x^2 \xi_x^{-2} \approx n_x M_x \qquad (2)$$

where ξ_x is the relevant correlation length of the phase transition and

$$n_x \approx \xi_x^{-3} \qquad (3)$$

is the number density of disconnected string loops, which one would expect to be formed with typical length of the order of the minimum value ξ_x, and hence with typical loop mass given in order of magnitude by

$$M_x \approx m_x^2 \xi_x, \qquad (4)$$

both the mass-energy per unit length U and the tension T of the string being expected to have comparable order of magnitude given by

$$U \approx T \approx m_x^2. \qquad (5)$$

A relatively small number of loops with exceptionally high length $L \gg \xi_x$ would also be expected to be formed, both during the phase transtion and later on as the result of self intersections of infinite strings whose initial contribution to the collective mass density of the network would be expected to be of the same order as the loop contribution[11]. Such exceptionally long loops are the only ones that survive to much later times in the standard non-conducting cosmic string scenarios[10] in which the initially numerous small loops are predicted to decay to nothing by gravitational radiation.

In the case of conducting string scenarios there is a major qualitative difference that was entirely overlooked in earlier discussions[12][13], having been first recognised only comparitively recently by Davis and Shellard[1]. The essential new feature is that the conducting string loops are characterised by a pair of independently conserved phase and charge quantum numbers N and C say, whose values, provided they are both non-zero, can be expected under very general conditions[1][2][3] to determine a corresponding classically stable ground state in the form of a rotating ring with a non-zero minimum mass-energy $M(C, N)$. This means that while some of the string loops present at later times will still belong to the minority formed during the phase transition with length $L \gg \xi_x$, or else will be loops formed later on by intersections, the numerically dominant contribution at much later times characterised by a cosmological temperature $\Theta \ll \Theta_x$ can be expected in the conducting case to consist of loops belonging to the initially dominant majority with mass density

$$\rho \approx M n \qquad (6)$$

where M is a characteristic mean mass of the individual ring states that are ultimately formed, and n is their number density whose magnitude may be estimated directly from (3) by application of the appropriate redshift factor, giving

$$n \approx (\Theta/\Theta_x)^3 n_x. \qquad (7)$$

The question that is of interest in the context of the dark matter problem is to compare the cosmic ring density ρ with the cosmological closure density ρ_c say at the same temperature Θ, which will be given by

$$\rho_c \approx m_c \Theta^3 \tag{8}$$

where the mass scale m_c is defined as the mass per black body radiation photon that would be required for cosmological closure, which is which is observationally known to be given in rough order of magnitude by $m_c \approx 10^{-26}$ ($\approx 10^{-2}$ e.v.).

In order to evaluate the all important closure ratio

$$\frac{\rho}{\rho_c} \approx \frac{M}{m_c}(\xi_x \Theta_x)^{-3} \tag{9}$$

(which can certanly not be substantially in excess of unity, in order for us to have escaped premature collapse of the universe) let us start by considering the conditions determining the likely value of the typical cosmic ring mass M.

For a single isolated cosmic ring of radius r the mass is given[2][3] in terms of the energy per unit length (in the locally preferred corrotating frame) U and the corresponding tension T (which will be given as a function of U by a model dependent equation of state) by the expression

$$M = 2\pi r(U + T), \tag{10}$$

the corresponding angular momentum value, J, being given by

$$J = 2\pi r^2 (UT)^{1/2}. \tag{11}$$

The derivation of these equations is based on the assumption that the gravitational and electromagnetic self interaction of the ring on itself is negligible compared with the the direct mechanical string tension and centrifugal forces, which will be true provided we have

$$U \ll 1 \tag{12}$$

and

$$I^2 \ll T \tag{13}$$

where I is the magnitude of the electromagnetic current in the string. In view of the relation (5) between the magnitude of the mass energy per unit length U and the symmetry breaking mass scale m_c it follows from the presumption that the latter is small compared with the Plank mass, so that the corresponding gravitational coupling constant satisfies

$$m_x^2 \ll 1 , \tag{14}$$

(since the largest value that one might reasonably wish to consider is the "grand unification" value $m_x^2 \approx 10^{-6}$) that there will never be any danger of violating the condition (12) for negligibility of self gravitation. It also seems safe to assume that (13) will be satisfied except in rather exceptional circumstances in view of the prediction[6][12] that the current in a superconducting string model should never exceed a maximum given by

$$\mu^2 \underset{\sim}{<} m_x^2 \tag{15}$$

where

$$\mu = I/e \qquad (16)$$

is the quantity identifiable[3][4][5] in the "magnetic" case (i.e. when the current in the ring is space-like) as the effective mass per unit phase winding number, as defined in terms of the corresponding winding number per unit length ν say by

$$\mu = dU/d\nu \ , \qquad (17)$$

while e^2 is the relevant charge coupling constant, which in the electromagentic case will have the moderately small value $e^2 \approx 1/137$. In a "magnetic" ring equilibrium state, these quantities determine the ratio of the conserved charge and phase winding quantum numbers by a relation[2][3] expressible as

$$\frac{C}{N} = 2\pi\kappa c_\tau \qquad (18)$$

where the condensate function κ and the characteristic speed of transverse perturbations c_τ are slowly varying functions of state given by

$$\kappa = \frac{\mu}{\nu} \ , \quad c_\tau^{\ 2} = T/U \ , \qquad (19)$$

with order of magnitude given by

$$\kappa \approx \kappa_0 \ , \quad c_\tau^{\ 2} \approx 1, \qquad (20)$$

where κ_0 is the value of κ in the zero currents limit, a quantity that is arguably the most important single parameter characterising the qualitative behaviour of the string at a macroscopic level. Although highly model dependent, the parameter κ_0 might be typically expected to differ from unity by at most a few orders of magnitude on dimensional grounds. So long as the magnitude of the string tension T actually is comparable to the mass-energy per unit length U (to which it would be exactly equal in the non-conducting case) in accordance with (5), the requirement (13) will obviously hold automatically as a consequence of (15).

Several recent authors[14][15] have demonstrated that in practice the condition (15) for the current to remain confined within the cross sectional radius $\approx m_x^{-1}$ of the string is likely to be unduly generous, the maximum current that can be carried in practice being more severely limited so that the rough inequality symbol $\lesssim$ could most often be replaced by a strong inequality $<<$ in (15). Such "current saturation" effects will also[2] give rise to lower limits on the value of the tension T. It is nevertheless possible, albeit at the expense of considerable ingenuity, to contrive particular superconducting string models in which the string tension can actually go to zero, and for which as well as the normal cosmic ring type equilibrium states, it would also be possible to construct an alternative *strictly static* (non-rotating and not necessarily circular) kind of equilibrium state[16][17], in which the condition (13) would of course be violated. (Such exceptionally occurring static zero tension states have sometimes been referred to as "cosmic springs", a term which is rather misleading in so far as the concept of a "spring" is normally to be understood as implying the existence of a restoring force tending to preserve a preferred configurations, whereas zero tension string states are effectively "floppy" with no such restoring force, since any negative tension must involve absolute instability [3][4] except in the extreme case of a loop so small that the length L of its circumference is comparable with the string cross sectional radius $\approx m_x^{-1}$ so that its geometrical configuration would be that of a torus so thick that description as a string would be inappropriate.) If they occur at all, such zero tension static ("floppy" rather than "springy") string states might

conceivably be of cosmological interest, but the likelihood of this is very small compared with that for the finite tension stationary (as opposed to strictly static) cosmic ring states which would appear to exist as a generic phenomenon for *all* kinds of superconducting string model.

In order to calculate the exact numerical value of the ring mass M as given[2][3] by (10), one not only needs to know the values of the conserved phase and charge numbers N and $C = Q/e$ (where Q is the total electric charge of the ring) by which it is specified, but one also needs to know the equation of state giving U as a function of T (or of ν) in the model. However in the case of the angular momentum J the situation is simpler, since *independently* of the equation of state one just has

$$J = CN. \tag{21}$$

Without knowledge of the precise equation of state that is relevant, the best one can do is to use the order of magnitude estimates (5) to evaluate the order of magnitude of the ring radius very crudely from (10) as

$$r \approx N^{1/2} C^{1/2} m_x^{-1} \tag{22}$$

and hence by (9) to obtain a corresponding crude estimate of the ring mass as

$$M \approx N^{1/2} C^{1/2} m_x \tag{23}$$

Having got to this point, it remains to form some opinion of the likely values of the quantum numbers N and C themselves, which, if we assume that subsequent loop intersections are not too important, will be given directly by their values during the phase transition whereby the string loops are formed in the first place. For a provisional estimate, it would seem reasonable, on the basis of a random walk argument, to guess that the phase numbers characterising a typical loop would come out to be given very roughly by

$$N \approx (\xi_x/\lambda_x)^{1/2} \tag{24}$$

where λ_x is the phase correlation length of the charge carriers and ξ_x is the presumably much longer vacuum defect correlation length as already introduced above; on the supposition that typical charge carrier phase speeds are comparable with unity one expcts the conserved number ratio to be of the order of the critical value[2][3]

$$\frac{C}{N} \approx 2\pi\kappa_0 \ . \tag{25}$$

We are thus tentatively lead to the crude estimates

$$r \approx \left(\frac{\kappa_0 \xi_x}{\lambda_x}\right)^{1/2} m_x^{-1} \tag{26}$$

and

$$M \approx \left(\frac{\kappa_0 \xi_x}{\lambda_x}\right)^{1/2} m_x \ , \tag{27}$$

which gves

$$\frac{\rho}{\rho_c} \approx \frac{m_x}{m_c} \kappa_0^{1/2} \lambda_x^{-1/2} \xi_x^{-5/2} \Theta_x^{-3} \tag{28}$$

In the standard non-conducting scenario one expects in the long run to approach a sparse string loop distribution that evolves in a self similar manner that is insensitive to the value of the relevant correlation length ξ_x during the phase transition, which in any case is restricted by

causality to be less than the value R_x of the Hubble radius at the time of the transition, which may be estimated as

$$R_x \approx \Theta_x^{-2} \qquad (29)$$

on the assumption that at thermal radiation dominates the dynamics of the universe at that stage, this value of R_x being commonly used as a conservative estimate for ξ_x. However in a conducting scenario the conservation of individual loops in the form of cosmic rings ensures that there will be no such tendency to self-similarity, the ensuing distribution remaining at all stages very sensitive to the value of ξ_x as can be seen from its explicit appeareance in (28). It will therefore be of importance to know whether ξ_x is in fact comparable with the causal upper limit (29) or whether it will be nearer the thermal fluctuation lengthscale which can be used as a first guess for the relevant correlation length for the charge carrying field,

$$\lambda_x \approx \Theta_x^{-1} . \qquad (30)$$

This lengthscale (30) would also be appropriate as an estimate for ξ_x if the phase transition were sufficiently rapid, and will obviously give a lower limit, the correct value lying somewhere in the range

$$\lambda_x < \xi_x < R_x. \qquad (31)$$

Taking account of the fact that the relevant cooling timescale is given directly by R_x itself, dimensional considerations would suggest that an intermediate value given by a power law formula of the form

$$\xi_x{}^3 \approx \lambda_x{}^{(3-\alpha)} R_x{}^\alpha \qquad (32)$$

is likely to be applicable, with an index value somewhere in the range $0 < \alpha < 3$, provided that the relevant coupling constants are not too far from unity. The plausibility of such a formula is confirmed by a rough estimate of Kibble[9] which, moreover, would appear to indicate that the appropriate index value is given simply by $\alpha = 1$.

Proceeding on this basis one obtains the very rough estimate

$$\xi_x \approx m_x^{-(1+\alpha/3)}, \qquad (33)$$

with

$$\lambda_x \approx m_x^{-1}. \qquad (34)$$

Since such a crude dimensional estimate can hardly be expected to be accurate to within better than a couple of powers of ten, it will not be very significantly different from the usual causal upper limit estimate $\xi_x \approx m_x^{-2}$ if the phase transition under consideration is that of "grand unification" for which the appropriate mass scale is given by $m_x \approx 10^{-3}$ ($\approx 10^{-16}$ G.e.v.). However if the relevant phase transition occurs at a much lower energy level, there will be an important distinction between the more plausible estimate based on $\alpha = 1$ and the usual crude causal estimate which corresponds to taking $\alpha = 3$.

In view of the accumulated loss of accuracy in the successive steps taken so far, any further loss of accuracy in assuming

$$\kappa_0 \approx 1 \qquad (35)$$

is unlikely to be important, so we shall adopt this further simplificaton from here on. Proceeding on this basis, it can be seen that, on substitution of the values tentatively estimated above, one obtains

as a first guess that the typical value of the charge and phase winding numbers in the product (21), determining the residual value of the angular momentum in the ultimate ring equilibrium state of the loop, will be given by

$$C \approx N \approx m_x^{-\alpha/6} \tag{36}$$

The corresponding expressions for the typical radius and mass of the ensuing cosmic ring states will thus be

$$r = m_x^{-1-\alpha/6} \tag{37}$$

and

$$M = m_x^{1-\alpha/6} \tag{38}$$

In order to give equilibrium configurations sufficiently extended to be considered as thin classical " rings" (as opposed to thick toroidal "vortons"[1]) the product CN (i.e. the angular momentum quantum number) must of course be large compared with unity. It is apparent from (35) that this condition will be only marginally satisfied if the relevant phase transition is that of "grand unification" with $m_x \approx 10^{-3}$, but that it will be satisfied by a large margin for conducting string producing phase transitions at the very much lower energies that are the most that would seem cosmologically allowable: it will be now be shown that a mass scale anywhere near as large as that required for grand unification would correspond to ring production so prolific that (like monopole production at a comparable energy scale) it would lead (as pointed out by Davis and Shellard[1]) to premature collapse of the universe, which means that the appropriate field theory must be such as to exclude the formation of such superconducting strings (and monopoles) or else that if formed they are somehow "inflated away" so that whichever way it is, the phase transition that would be relevant for the production of actually surviving cosmic rings (or monopoles) would be one occuring subsequently with a mass scale m_x having a considerably smaller value which we shall now attempt to evaluate.

Substituting the above estimates into the formula (28) for the ratio of the ring distribution mass density to the cosmological closure density gives

$$\frac{\rho}{\rho_c} \approx m_c^{-1} m_x^{1+5\alpha/6} \ . \tag{39}$$

This ratio may certainly not substantially exceed unity, but should not be very much smaller if the density of the cosmic ring distribution is to be cosmologically significant. The phase transition mass scale that is of interest from the point of view of these requirements is thus obtained as being given roughly (in Plank units) by

$$m_x \approx m_c^{6/(6+5\alpha)} \tag{40}$$

On the basis of the known numerical value $m_c \approx 10^{-26}$ this works out very roughly, for the Kibble value $\alpha = 1$, as $m_x \approx 10^{-15} \approx 10^4$ G.e.v.

This critical mass scale (representing an upper limit on what is consistent with the avoidance of cosmological mass excess in the formation of all except non-conducting cosmic string varieties) is not to be confused with the typical mass M that is expected to characterise the resulting cosmic rings themselves. Substitution of (40) in (38) gives for the typical mass of an idiviual ring the significantly higher estimate

$$M \approx m_c^{(6-\alpha)/(6+5\alpha)} \tag{41}$$

which works out very roughly, for $\alpha = 1$ as $M \approx 10^{-13} \approx 10^3$ T.e.v.

This expected typical mass thus obtained as a requirement for marginal cosmological closure is consistent, but only just, with the "ultra-massive" range that is required (to avoid the observable consequences of excessive number density) for the the kinds of "chump" dark matter candidate that has most commonly been considered in other physical contexts so far[7][8]. While the mass is in this sense large, the associated typical ring radius r would nevertheless be small compared with that of an ordinary atomic nucleus, being given by

$$r \approx m_c{}^{-(6+\alpha)/(6+5\alpha)} \tag{42}$$

which, for $\alpha = 1$ works out as something like 10^{-17} cm. This radius exceeds the microscopic cross sectional radius of the string defect itself by a comparatively modest factor of the order of magnitude of the winding number N and hence, on the basis of the assumption (36), also of the charge number C, whose expected order of magnitude is obtainable from the foregoing results as

$$C \approx N \approx m_c{}^{-\alpha/(6+5\alpha)} \tag{43}$$

which is only very weakly dependent on α, working out numerically for $\alpha = 1$ as $C \approx N \approx 10^2$. It is to be noticed, as another interesting (conceivably anthropic?) numerical coincidence, that the charge numbers to be expected on the basis of the foregoing (admittedly rather tenuous) chain of reasonning fall roughly within the same range as those of the familiar chemical elements, for whom the upper charge number cut off is of course of the order of the inverse fine structure constant, $1/e^2 \approx 137$ which is the highest value that can resist immediate discharge by pair creation.

References.

[1] R.L. Davis, E.P.S. Shellard, Physics Letters **B209**, 485 (1988).

[2] B. Carter in *Formation and Evolution of Cosmic Strings*, ed. G.W. Gibbons, S.W. Hawking, T. Vachaspati (Cambridge U.P. 1990).

[3] B. Carter, Physics Letters **B238**, 166 (1990).

[4] B. Carter, Physics Letters **B224**, 61 (1989).

[5] B. Carter, Physics Letters **B228**, 446 (1989).

[6] E. Witten, Nucl. Phys. **B249**, 557 (1985).

[7] H. Goldberg, in *Progress in Weak Interactions* (XXI st Moriond Meeting, 1986), ed J. Tran Thanh Van, Vol I, p467 (Frontières, Paris, 1987).

[8] A. De Rujula, S.L. Glashow, U. Sarid "Charged Dark Matter", preprint C.E.R.N. TH 5490 (1989).

[9] T.W.B. Kibble, Physics Reports, **67**, 183 (1980).

[10] A. Vilenkin, Physics Reports, **121**, 263 (1965).

[11] F.R. Bouchet, D.P. Bennet, in *Dark Matter* (XXIIIrd Moriond Meeting, 1988), ed. J. Audouze, J. Tran Thanh Van, 361 (Frontières, Paris, 1988).

[12] J.P. Ostriker, C. Thompson, E. Witten, Physics Letters **B180**, 231(1986).

[13] E.M. Chudnovsky, G.B. Field, D.N. Spergel, A. Vilenkin, Phys. Rev. **D34**, 944 (1986).

[14] C.T. Hill, H.M. Hodges, M.S. Turner, Phys. Rev. **D37**, 263 (1988).

[15] A. Babul, T. Piran, D.N. Spergel, Physics Letters **B202**, 307 (1988).

[16] E. Copeland, M. Hindmarsh, E. Turok, Phys. Rev. Letters **58**, 1910 (1987).

[17] D. Haws, M. Hindmarsh, N. Turok, Physics Letters **B209**, 255 (1988).

DARK MATTER AND WIMPS

SCINTILLATION DETECTORS FOR DARK MATTER SEARCH

P. Belli, R. Bernabei, S.d' Angelo
Dipartimento di Fisica
II Università di Roma
and INFN, Sez. di Roma II

A. Incicchitti, D. Prosperi
Dipartimento di Fisica
Università di Roma "La Sapienza"
and INFN, Sez. di Roma

(Presented by P. Belli)

Abstract

In this paper we point out the possibility to use NaI(Tl) and liquid xenon scintillators as target-detectors to search for WIMPs.

1. Introduction

Due to the strong evidence[1,2,3] that the "visible" matter in the Universe is about two orders of magnitude lower than expected by the dynamical analysis of astrophysical objects, efforts are devoted to verify the "dark matter" candidate solution. In particular, it was pointed out that this missing mass could be justified by the existence of Weakly Interacting Massive Particles [WIMPs(*)][4] and experiments are in progress[5,6] to detect candidates that interact by elastic scattering on various target nuclei[7,8,9]. In this case the recoil kinetic energy of the nucleus is the measurable quantity. This recoil energy could be evaluated by the elastic scattering kinematics[10] of a WIMP - with M_w mass and v velocity - impinging on a M mass nucleus as: $T = E_1 \left(\dfrac{1 - \cos\theta^*}{2} \right)$,

with $r = \dfrac{4 \cdot M \cdot M_w}{(M + M_w)^2}$, $E_1 = \dfrac{1}{2} M_w v^2$ and θ^* the c.m. scattering angle. If the WIMP velocity is equal to the virial one in our Galaxy ($v \approx 10^{-3}c$) and $M_w \approx 10\text{-}200$ GeV, the detectable energy is about $10 \div 100$ keV. Furthermore, assuming an isotropic c.m. angular distribution and a Maxwellian velocity distribution (with $v_0 = 220$ km/s parameter), the energy spectrum is:

$\dfrac{dR}{dT} = \dfrac{R_0}{E_0 r} \exp\left\{ - \dfrac{T}{E_0 r} \right\}$, with $E_0 = \dfrac{1}{2} M_w v_0^2$ and R_0 equal to the total

number of observed events.

(*) Notice that the LEP experiments has excluded a forth neutrino family and agree with the "dark matter" silicium experiments in excluding some supersymmetric new particles (possible candidates also as d.m. particles),with masses of some tenth GeV. For this reason, also in the search of d.m. candidates the interest is now to the detection of higher mass particles, i.e. to heavier target nuclei (like those we propose in the following).

The "dark matter" particles could interact either by weak coherent v-like vector currents or by spin-dependent strength[9]. The particles interacting through weak coherent v-like vector currents, are elastically scattered from nuclei by Z^0 exchange. In the non relativistic limit, the counting rate can be obtained by the following cross section[9]:

$$\sigma = \left(2.13 \times 10^{-35} \frac{cm^2}{1 \ GeV^2}\right) \frac{4M_w^2 M^2}{(M_w+M)^2} \ \overline{Y}^2 \left(\frac{\overline{N}}{100}\right)^2$$

where $\overline{Y}$ is the mean value of the weak hypercharge, $\overline{N}$ is equal to $N-(1-4\sin^2\theta_w)Z$ with N and Z numbers of neutrons and protons in the target nucleus and θ_w is the Weinberg angle. This formula does not take into account the finite size of the nucleus that causes loss of coherence at high transferred momenta, so that a correction factor is needed as pointed out in ref. 11.

In case these particles interact by spin-dependent strength, the cross section is proportional to $J(J+1)$ where J is the nuclear spin. Then, being $J = 0$ for even-even nuclei in the ground state, any even-even nuclear state cannot allow to detect axially coupled particles. However, notice that the cross section values estimated in the literature are, in this case, lesser than in the previous one[9].

In table 1 we summarize the expected counting rate on some target-nuclei[9]. As discussed in the following, we are interested in using NaI(Tl) and liquid Xenon, looking to the scintillation due to the recoil nucleus[12,13].

Table 1

A short review of typical orders of magnitude for the interaction rates [9].

Interactions with nuclei	candidate particles	target nuclei	typical rates (kg^{-1} day^{-1})
spin-independent: coherent weak interaction	sneutrino; heavy v Dirac	Al	150 (M_δ=10 GeV)
		Ge	700 (M_δ=10 GeV)
		Sn	1400 (M_δ=10 GeV)
		Al	120 (M_δ=100 GeV)
		Ge	1400 (M_δ=100 GeV)
		Sn	4700 (M_δ=100 GeV)
spin-dependent	photino; z-ino; higgsino; heavy Majorana v	Hg	5 (M_δ=5 GeV)
		Ge	1.5 (M_δ=5 GeV)
		F	1 (M_δ=5 GeV)
		Hg	1 (M_δ=17 GeV)
		Ge	0.3 (M_δ=17 GeV)
		F	0.4 (M_δ=17 GeV)

Notice that most of the existing detectors for "dark matter" are ionizing detectors built for other underground experiments like double beta decay: i.e. germanium or silicium detectors[8] sensitive to recoil nuclei with energy greater than ~12 keV. The typical background rate in these experiments is: ~1 event kg^{-1} keV^{-1} day^{-1}. Having zero spin most of the isotopes in germanium and silicium, they cannot detect axially coupled particles; furthermore, some technical difficulties do not allow to build large dimension detectors.

Because the counting rate for WIMPs and target-nuclei interactions is low, the most relevant problem is the background rejection using low activity materials and installing the experiments deep underground. It is clearly important to use a large amount of target material.

Furthermore, a "dark matter" signal can be discriminated from the background looking to: i) the annual modulation of

the counting rate and of the energy spectrum; ii) the recoil nucleus direction (but not in the cases of scintillators); iii) the energy spectrum shape, that gives the E_{or} value and an estimate of the mass of the "dark matter" particle; iv) the using of various target-materials and ,then, the different cross sections for spin-dependent or spin-independent interactions. In particular, working with scintillation detectors the data could be analyzed looking to iii); i.e. the experimental energy spectrum is compared with those expected for various M_W and the agreement is verified by best-fit methods[8]. An important confirmation could arise looking for i) modulation.

2. Scintillators as target-nuclei

The scintillators could be interesting materials for "dark matter" detection, being cheap and easy to build with large size, having good purity, a controllable light response and, finally, opening a new set of target-nuclei (that in principle could allow to explore also the axial coupling). In the following, we will devote our attention to experiments with low activity NaI(Tl) or liquid Xenon targets, where the scintillation pulse by the recoil nucleus could be detected.

In fig. 1 the behaviour of the nucleus mean recoil energy as a function of M_W in case of these materials is shown and compared with those for germanium, silicium and hydrogen (low pressure TPC). This figure clearly shows that the NaI(Tl) - operating with a few keV energy threshold - could allow to detect "dark matter" particles with masses higher than 10 GeV, while the liquid Xenon become competitive when detecting candidates with masses higher than 100 GeV.
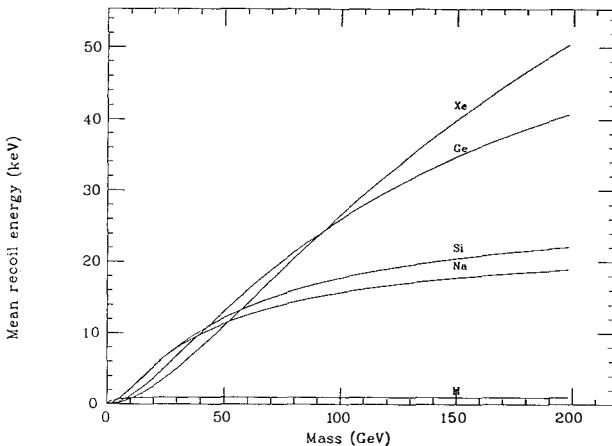
Fig. 1 - Mean target-nucleus recoil energy as a function of the "dark matter" particle mass.

2.1 The NaI(Tl)

The NaI(Tl) has a high light response (that could allow a high signal to noise ratio) and is able to reach a few keV energy threshold[13]. The pulse height (i.e. the energy) due to the ^{23}Na recoil nucleus after elastic scattering with a 200 keV neutron has been already measured[13] to evaluate (comparing the detected with the kinematic value) the relative ionizing efficiency; a value of 0.38 has been found.

To know how far the commercial large NaI(Tl) detectors are from the requirements of a "dark matter" experiment, we used a 3 Kg NaI(Tl) seen by an EMI9839B photomultiplier. This detector was placed inside a 5 cm thick lead shield in the Gran Sasso underground laboratory ($\approx$ 3500 m.w.e.) for a short period. In fig. 2a), b) and c) we show the response of this counter to a ^{137}Cs source (i.e. $\approx$ 35 KeV X rays), to a ^{57}Co source (i.e. $\approx$130 KeV photons) and the underground counting distribution. Because the ratio between high ionizing particles and photons light outputs is about 0.5 in NaI(Tl), we

can assume that the response of a specially built crystal could be good at least in the 30÷200 KeV energy region.

A)

B)

C)
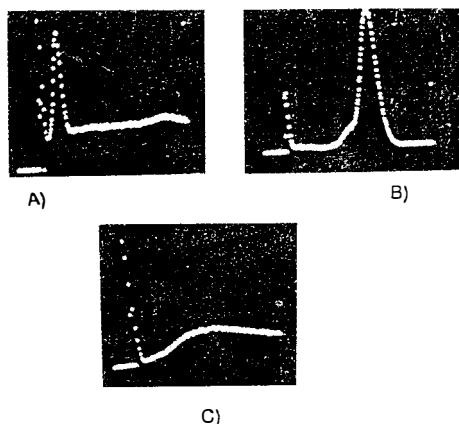
Fig. 2- a) Counter response to ≈ 35 KeV X rays from ^{137}Cs source; b) counter response to ≈ 130 KeV photons from ^{57}Co source; c) counting distribution deep underground.

In table 2 we report the expected rate for the "dark matter" candidates we are interested in. It is clear from this table that we need to use special low activity crystals and photomultipliers, that will allow to obtain a background rate ≤ 5 cpd/Kg/keV (in this case, e.g. we could measure recoil energy between 30 and 100 keV due to spin-independent interaction of a 100 GeV mass "dark matter" particle).

Notice that NaI(Tl) could allow - in principle - to detect also axially coupled particles, being the spin of the fundamental state of ^{23}Na $\frac{3^+}{2}$ and that of ^{127}I $\frac{5^+}{2}$; but, we think extremely hard to reach the background level needed to measure such processes.

Table 2

Measured "background" in a 3 Kg NaI(TI) compared - as an example - with the expected rates for 100 GeV mass spin-independent interacting "dark matter" (the coherence factor was already taken into account)

energy range (keV)	expected rates for 100 GeV d.m. particles (cpd/kg/keV)	measured background (cpd/kg/keV)
30-50	25.6	346
50-100	7.8.	778
100-200	0.70	864
200-300	0.02	691
300-400	-	605
400-500	-	518

2.2 The liquid Xenon.

We have already pointed out[12) the interest to use the liquid xenon as target material because of its scintillation properties; in fact, as in the NaI(TI) case, we think that it could be possible to detect the light pulse due to the recoil nucleus.

The linearity, energy resolution and threshold value for this kind of detectors are discussed in ref. 14; from the data quoted therein one could think to work with a threshold energy of about 15-20 keV and an energy resolution of about 12% at 122 keV. Furthermore, we remarks the high ratio between the α and electrons pulse heights: (1.1±0.2) and the low intrinsec background (radioactive nuclides ≤0.1 ppm); however, we need to know the light response of the detector

to the xenon recoil nuclei, i.e. to measure the relative ionization efficiency at a neutron beam.

Now we are working on a 118 cc (i.e. about 61 liters of gaseous xenon NTP) prototype[15]. Because the wavelenght of the scintillation light is in the far UV region, an efficient light collection is a real problem; however, in our prototype we collect this UV emission using UV photomultipliers decoupled in vacuum from the detector UV windows. The gas we utilize has 99.998 % purity increased by a further purification with oxisorb cartridge and a suitable high temperature getter.

In fig. 3 the pulse shape recorded by a transient digitizer LECROY 6880A from a [22]Na source as seen by our detector is shown.
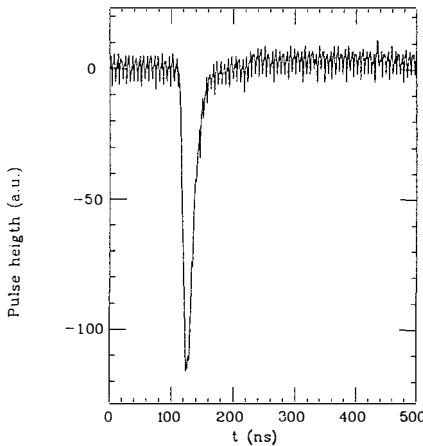
Fig. 3- Pulse shape recorded by a transient digitizer LECROY 6880A from a [22]Na source.

In fig. 4a) the expected counting rate as a function of M_W and with various energy thresholds is shown, while in fig. 4b) the expected counting rate as a function of the energy released by the recoil nucleus, e.g. for $M_W = 50$ GeV in three different periods of the year (i.e. with mean velocities: 240,

270 e 300 km/s) is reported. As mentioned above, the presence of such an annual modulation could contribute to the background rejection.
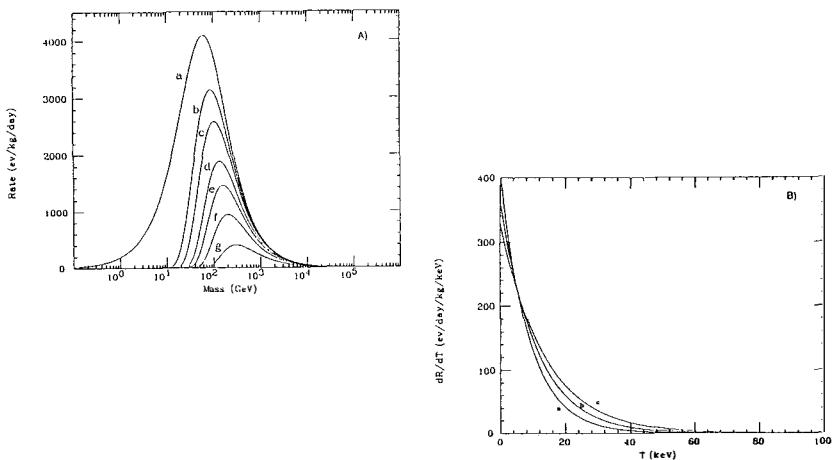
Fig. 4 - A) The expected counting rate for "dark matter" detection. in liquid xenon (corrected for loss of coherence) as a function of the particle mass in case of various detection energy thresholds: a) no threshold; b) 5 keV; c) 10 keV; d) 20 keV; e) 30 keV; f) 50 keV; g) 100 keV. B) The expected counting rate for "dark matter" detection in liquid xenon (corrected for loss of coherence) as a function of the energy released by the recoil nucleus in case of a 50 GeV impinging particle during three different periods of the year (i.e. mean velocities: a) 240, b) 270 and c) 300 km/s).

In table 3 we summarize the counting rate as a function of the particle mass, M_W, in case of spin-independent interactions (with same hypotheses than before). As one can see, the expected counting rate in xenon is promising when compared with other materials[8]. As an example, with a threshold of about 20 keV, the counting rate on a 118 cc detector is about 610 ev/d for 100 GeV mass particles.

Table 3

Calculated counting rates (already corrected by coherence factor) for "dark matter" spin-independent candidates in a liquid xenon target-detector.

M_W (GeV)	Calculated counting rates (ev/Kg/day)
10^1	1650
10^2	3756
10^3	571
10^4	56
10^5	6

Finally, regarding the possibility to detect axially coupled particles, notice that the natural xenon contains the following isotopes with non zero spin: ^{129}Xe (26.4%, spin 1/2+) and ^{131}Xe(21.2%, spin 3/2+)

3. Conclusions.

In conclusion, we think that low activity NaI(Tl) and liquid xenon scintillation target-detectors could be very suitable to contribute to the search for "dark matter" candidates. They could allow to rule out complementary regions in the σ-M_w plane and to state new M_W upper limits.

Referenze

1-Zwicky F., Helv. Phys. Acta, **6** (1933) 110; Oort J.H., Bull. Astr. Instr. Neth. , **6** (1932) 249.
2-Rubin V.C. et al., Ap. J. **238** (1980) 471; Bosma A., Ap. J. **86** (1981) 1825.
3-Trimble V., Contemp. Phys. **19** (1988) 373.

4-R.L. Gilliland et al., Ap. J. 306 (1986) 703; W.H. Press and D.N. Spergel, Ap. J. 296 (1985) 679.

5-G.F. Giudice, FERMILAB-CONF-89/15-T} (1989); V. Trimble, Contemp. Phys. 19 (1988) 373.

6-P.F. Smith, J.D. Lewin, RAL-88-045} (1988); E. Fiorini, T.O. Niinikoski, Nucl. Instr. and Meth. 224 (1984) 83; L. Gonzales-Mestres and D. Perret-Gallix, pre-print LAPP- EXP-88-18} (1988); L. Gonzales-Mestres and D. Perret- Gallix, pre-print LAPP-EXP-89-07} (1989).

7-K. Griest, Phys. Rev. Lett. 61 (1988)666; A.K. Drukier, K. Freese, D.N. Spergel, Phys. Rev. D33 (1986)3495; G.D. Coughlan, Z.A. Lalak, Phys. Lett. B199 (1987)275.

8-D.O. Caldwell et al., Phys. Rev. Lett. 61 (1988)510; S.P. Ahlen et al., Phys. Lett. B195 (1987)603.

9-Goodman M.W. and Witten E., Phys. Rev., **D31** (1985) 3059.

10-Tao C. " Wimps: a short review" , DPhPE 89-02.

11-D. Z. Freedman, Phys. Rev., **D9** (1974) 1389.

12-Belli P. et al., "Detecting "Dark Matter" using a liquid xenon scintillation detector", to be published on Nuovo Cim.

13-Gerbier, "The Saclay program for dark matter search", to appear on Proceed. of Moriond Workshop on new and exotic phenomena, Les Arcs, jan 1990.

14-Barabanov I.R.,Gavrin V.N. and Pshukov A.M., Nucl. Instr. & Meth., **A254** (1987) 355.

15-P. Belli et al., int. report **ROM2F/022/89** (1989).

A BOLOMETRIC APPROACH
TO COSMION SEARCHES

A. de Bellefon, P. Espigat

L. P. C. - Collège de France, Paris

L. Gonzalez-Mestres

L. A. P. P. Annecy

Abstract

Recent progress in bolometric techniques for astronomy and particle physics suggests the possibility to built composite bolometers with absorber masses in the range 100 mg - 1 g , operating at 1 keV threshold and $T \simeq 30$ mK . We propose to build and test such bolometers in view of cosmion searches: with appropriate targets (100 mg of pure Si, C,... for cosmions with vector interaction; 1 g of F, Li... in the crystal for cosmions with axial interaction), it would be possible to reach ~ 1 event/day . Background rejection and astronomical reliability (threshold in galactic speed) would thus be considerably improved as compared to the potentialities of any ionization detector. We particularly discuss a possible search for cosmions with vector interaction, using a 100 mg silicon bolometer with a 1 keV threshold.

1. ASTROPHYSICAL ASPECTS AND EVENT RATES

In spite of the cosmological significance of recent LEP results [1], it does not seem that the cosmion model can be ruled out by these data. Particles coupled to hadrons, but not to the electroweak and leptonic sectors, remain allowed. UA1, UA2 and CDF data [2] can also be used, but forces weak enough and coupled only to the hadronic sector would escape observation due to the severe background from strong interactions. The astrophysical search for particles characterized by their cross sections with protons and nuclei (able to solve the solar neutrino problem and being in position to provide the galactic dark matter) remains therefore well motivated. Existing experiments [3] are based on intrinsic semiconductor silicon, with a threshold around 1.5 keV in ionization, and a threshold in nucleus recoil of at best 5 keV according to Lindhard's theory [4] and recent calibrations [5]. Technical improvements may eventually yield a threshold of 0.6 keV in ionization, equivalent to 3 keV in recoil. Even so, a cosmion of mass $m = 2$ GeV depositing 3 keV in a silicon detector must have a speed with respect to earth of at least 1040 Km/s, which after subtraction of the earth galactic velocity (about 240 Km/s), leads to a threshold of 800 Km/s in cosmion galactic speed. This is obviously too high, as compared to the escape velocity, usually estimated to 600 Km/s .

A threshold at 1 keV would be a natural way out, leading for $m = 2$ GeV to a threshold in cosmion galactic speed around 360 Km/s , which is acceptable on astronomical grounds (the average velocity, v_{rms} , is often taken to be $\simeq 270$ Km/s). Such a threshold in nucleus recoil cannot be reached by bulk semiconductor detectors. The expected event rate would increase considerably by working at 1 keV threshold, and may therefore lead to a better background rejection. Assuming a maxwellian distribution of cosmion speed in the halo, and using the notations of [6], the differential event rate dN/dE_R (E_R = recoil energy) is given by:

$$dN/dE_R \;=\; \pi^{1/2}/4 \; N_0 \; (E_0 \, r \, y)^{-1} \; [erf(x+y) \;-\; erf(x-y)] \tag{1}$$

where N_0 is the event rate per unit of detector mass:

$$N_0 \;=\; 2\pi^{-1/2} \; \rho/m \; \beta_0 \, c \; \sigma \; N_A/A \tag{2}$$

ρ = galactic cosmion mass density $\simeq 0.4$ GeV cm^{-3} , $\beta_0 \, c = (3/2)^{-1/2} \, v_{rms}$, σ = cosmion-nucleus cross section , N_A = Avogadro number, A = molar mass , $E_0 = 1/2 \, m \, \beta_0^2$, $r = 4mM(m+M)^{-2}$, M = nucleus mass, $x = (E_R/E_0 r)^{1/2}$, $y = \beta_e/\beta_0$, $\beta_e \, c$ = earth galactic speed and:

$$erf(z) = 2\pi^{-1/2} \int_0^z exp(-t^2) \, dt \tag{3}$$

The integrated flux $N(E_{th})$ at $E_R > E_{th}$ (threshold in recoil energy) is found to be: $N(E_{th}) = \pi^{1/2}/2 \; N_0 \; \beta_0/\beta_e \; \Sigma(E_{th})$, where:

$$\Sigma(E_{th}) = [z_+ z_-/2 - 1/4] \, [erf(z_-) - erf(z_+)] \;+\; (4\pi)^{-1/2} \, [z_+ \, exp(-z_-^2) - z_- \, exp(-z_+^2)] \tag{4}$$

where: $z_+ = x_{th} + y$, $z_- = x_{th} - y$, and $x_{th} = (E_{th}/E_0 r)^{1/2}$. The cross-section in (2) can in principle be computed form the cosmion cross section with single protons and neutrons. For cosmions with vector interaction, an isospin-independent interaction is often assumed, leading for the total event rate $N(E_{th} = 0)$ to a figure of merit:

$$\lambda \;=\; \sigma_H(m) \; (m+1)^2 M^3 m^{-1} (M+m)^{-2} \tag{5}$$

where all masses are given in hydrogen mass units and σ_H stands for the cross section with hydrogen. If the cosmion interaction were coupled only to protons, the cross section with a nucleus would be lower by a factor ≈ 4 .

The situation is much less clear for the case of axial interaction, which leads to spin-spin coupling in the non-relativistic limit. Even assuming that neutrons are coupled to the cosmion interaction, the theoretical and experimental ambiguities on the nucleon spin structure make it difficult to give precise predictions for nuclei were the spin would not be carried by protons. The safest approach, in this case, would be to use odd-even nuclei such as ^7Li , ^9Be , ^{11}B , ^{19}F , ^{27}Al , ... for which theoretical estimates of photino cross-sections exist in the litterature [7-9]. The figure of merit can then be expressed as:

$$\lambda = \sigma\,(m\,M)^{-1} \tag{6}$$

Typical values of masses, hydrogen cross sections, figure of merit for several target nuclei and event rates at $E_{th} = 1$ keV , can be found in Table 1 for cosmions with vector coupling to matter, and in Table II for the case of axial coupling. Heavier nuclei are not considered, as they lead to very small values of $E_0\,r$.

m (GeV)	σ_H (pb)	Target T	σ_T (nb)	λ (pb GeV^{-1})	N_0 (g^{-1} day^{-1})	$E_0 r$ (keV)	N/N_0 $E > 1$ keV	Rate at $E > 1$ keV (g^{-1} day^{-1})
2	20	^{7}Li	5	381	196	0.37	0.28	55
2	20	^{12}C	19	793	407	0.26.	0.14	56
2	20	^{28}Si	123	2195	1125	0.13	0.008	9
3	8	^{7}Li	3	146	75	0.67	0.59	44
3	8	^{12}C	12	327	168	0.51	0.46	77
3	8	^{28}Si	82	976	500	0.28	0.16	80
4	1	^{7}Li	0.5	18	9	0.99	0.79	7
4	1	^{12}C	2	42	22	0.80	0.68	15
4	1	^{28}Si	15	134	69	0.47	0.40	28
5	0.6	^{7}Li	0.36	10	5	1.31	0.89	4
5	0.6	^{12}C	1.6	26	13	1.12	0.83	11
5	0.6	^{28}Si	12	87	45	0.69	0.61	27
6	0.5	^{7}Li	0.35	8.3	4.2	1.60	0.96	4
6	0.5	^{12}C	1.6	22	11	1.43	0.92	10
6	0.5	^{28}Si	13	78	40	0.94	0.75	30
7	0.4	^{7}Li	0.3	6.4	3.3	1.88	1.02	3
7	0.4	^{12}C	1.5	18	9	1.75	0.99	9
7	0.4	^{28}Si	13	66	34	1.20	0.86	29

Table I : Typical event rates for cosmions with vector interaction at $E_{th} = 1$ keV .

σ_H (pb)	m (GeV)	^{1}H	^{7}Li	^{9}Be	^{11}B	^{19}F	^{27}Al
500	2	52	27	10	4	3	0.3
220	3	20	20	11	5	7	2.5
120	4	9	12	7	3	7	2.5
4.5	4	0.4	0.5	0.3	0.1	0.3	0.1
45	5	3	4	3	1.5	3	1.5
11	5	0.6	1	0.7	0.3	1	0.4

Table II : Approximate event rates (g^{-1} day^{-1}) for cosmions with axial interaction, using different targets and assuming E_{th} = 1 keV . Hydrogen cross-sections are taken from [10] and values of λ inferrred from [7-9]. Extrapolation from EMC data is used for ^{9}Be .

2. CHOICE OF DETECTOR AND STRATEGY

For cosmions with vector interaction, a 100 mg silicon detector with a 1 keV threshold appears as the best starting point. Silicon is easy to obtain in ultrapure form, and the technology to produce high quality crystals exists as well. Even with a background 10 times worse than that of the Oroville experiment, the exclusion plot of Fig. 1 would be obtained and bounds on cosmions crucially improved. One detection technique may provide the required performance with the present status of the art: indeed, *thermal composite bolometers* have produced in the last two years results [11] compatible with our goals.

Several papers on thermal bolometers and related techniques can be found in the Proceedings of the previous Moriond Astrophysics Workshop [12], and also in these Proceedings [13]. A thermal bolometer consists essentially of a single crystal (the absorber) on which a very sensitive thermometer (usually, a resistive one: the thermistor) is glued or implanted. In the ideal case, where the heat capacity of the detector is dominated by that of the absorber, its heat capacity follows a T^3 law and electronic as well as microphonic noise can be neglected (which is not the case for existing prototypes), the ultimate energy resolution would be given by thermal fluctuations and follows then the law:

$$\Delta E \quad \propto \quad M^{1/2} \ T^{5/2} \ \Theta_D^{-3/2} \tag{7}$$

The bolometer technique has undergone spectacular progress, demonstrating in particular the possibility to compensate an increase in detector mass by a decrease in operating temperature, according to (7). From $\Delta E_{FWHM} \simeq 36$ keV for 5 - 6 MeV α particles obtained in the early period [14] with a 0.25 mm^3 diamond bolometer at $T = 1.3$ K , it has been possible to reach a similar performance with a 280 g sapphire bolometer at 100 mK [15]. Particularly relevant to our problem are two results obtained in rather different conditions:

1. The Milan group reports [16] $\Delta E_{FWHM} \simeq 1\%$ on the 1.2 and 1.3 MeV γ rays of ^{60}Co with a 11 g germanium detector operated at $T \simeq 30$ mK . From the $M^{1/2}\Theta_D^{-3/2}$ law in (7) , one gets: $\Delta E_{FWHM} \simeq 500$ eV for 100 mg of silicon at the same temperature. This clearly suggests that a 1 keV threshold can indeed be obtained using similar techniques.
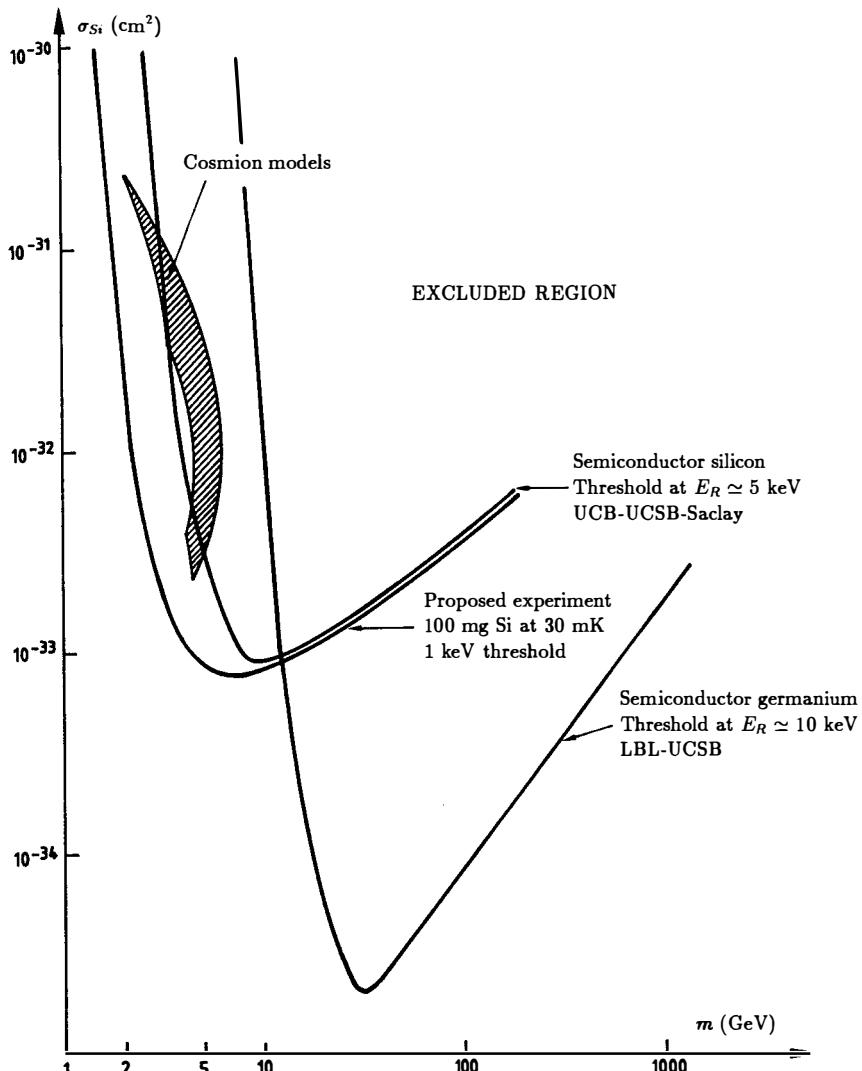
Fig. 1 - Exclusion plot (in terms of mass and silicon cross-section) for cosmions with vector interaction, from a 100 mg silicon bolometer working at 1 keV threshold and with a radioactive background of 1 evt g^{-1} keV^{-1} day^{-1}. Also shown are bounds from semiconductor silicon and germanium, obtained in experiments with much lower background rates.

2. Already in 1988, a group of astronomers [17] reported 80 eV rms noise with a small diamond bolometer operated at 150 mK , giving $\Delta E_{FWHM} \simeq 280$ eV on 5.9 keV ^{55}Fe γ's . The heat capacity, at 40 mK , of the thermometer used was 3×10^{-13} Joule K^{-1} , which is the equivalent of 20 mg of silicon at the same temperature. Again, it seems possible by a careful tuning of thermistor size and doping, to reach a 1 keV threshold with a 100 mg silicon bolometer operated at 30 mK . The use of NTD (neutron transmutation doped) thermistors [18] may be an interesting possibility, due to compensation and high doping uniformity.

It therefore seems urgent to build and test (down to $T \simeq 30$ mK) a 100 mg silicon composite bolometer incorporating a suitable sensor. If the 1 keV threshold is reached, such a detector installed at the Frejus laboratory will provide the first step of a long term dark matter search based on cryogenic techniques. A small ultrapure silicon bolometer would not only perform an efficient search for cosmions, but also measure for the first time the radioactive background at $E > 1$ keV , with unprecedented energy resolution. The use of silicon is probably safer than that of diamond, which in spite of excellent specific heat properties has to face isotopic (^{14}C) and chemical (for natural diamond) radioactive impurities. Next step can be the search for cosmions with axial coupling using a ~ 1 g LiF or boron absorber.

References

[1] See, for instance, L. Krauss, these Proceedings.

[2] Proceedings of the Moriond Session on Hadronic Interactions, March 1990.

[3] C. Tao, these Proceedings.

[4] J. Lindhard et al., K. Dan. Vidensk. Selsk. Mat.-Fys. Medd. 33, 10 (1963); 36, 10 (1968).

[5] G. Gerbier et al., Saclay Preprint DPhPE 90-02 .

[6] P.F. Smith and J.D. Lewin, Phys. Rep. 187, 203 (1990).

[7] J. Ellis and R. Flores, Nucl. Phys. B307, 883 (1988).

[8] A.F. Pacheco and D. Strottman, Phys. Rev. D40, 2131 (1989).

[9] J. Engel and P. Vogel, CALTECH preprint CALT-63-545 (1990).

[10] A. Bouquet, Ecole de Gif de Physique des Particules, September 1989.

For a detailed study of cosmion cross sections with vector and axial interaction, see: F. Martin de Volnay, these Proceedings.

[11] See, for instance, "Low Temperature Detectors for Neutrinos and Dark Matter - III", Frontières 1989.

[12] Proceedings of the Moriond Astrophysics Workhop, March 1989. Ed. Frontières.

[13] Y. Giraud-Héraud and C. Tao, these Proceedings.

[14] N. Coron et al., Nature 314, 75 (1985).

[15] W. Seidel et al. in [11].

[16] O. Cremonesi, Moriond Proceedings January 1990. Ed. Frontières.

[17] N. Coron et al. in "Superconductive and Low Temperature Particle Detectors", North-Holland 1988.

[18] E.E. Haller, IR Phys. 25, 257 (1985).

DARK MATTER CANDIDATES IN THE SUPERSYMMETRIC STANDARD MODEL

Keith A. Olive
School of Physics and Astronomy
University of Minnesota
Minneapolis, MN 55455

ABSTRACT

It is well known that the lightest particle in the supersymmetric standard model, the LSP, is stable, massive and can be a dark matter candidate. It has commonly been assumed to be the photino. However, the photino is the LSP only in a small portion of parameter space. In general, the LSP is a mixture of two gauginos and two Higgsinos. Nevertheless over a very large portion of the parameter space the LSP can still be described in terms of one of a few pure states, which are described here.

The need for substantial amounts of dark matter in the Universe is from both observational evidence and theoretical prejudice is well known and I will not review it here[1]. Instead I will concentrate on the candidates that are expected to arise in the supersymmetric standard model.

Supersymmetric theories introduce several possible candidates. If R-parity (which distinguishes between "normal" matter and the supersymmetric partners) is unbroken there is at least one supersymmetric particle which must be stable. For example, by defining $R = (-1)^{3B+F+L}$, all "normal" particles, γ, $W^{\pm}$, Z, H_1, H_2, e, μ, τ, ν's + quarks, antiquarks and gluons (there are two Higgs doublets in the minimal supersymmetric model, necessary to give both up quark and down quark masses) have R = +1. All of the supersymmetric partners, $\tilde{\gamma}$, $\tilde{W}^{\pm}$, $\tilde{Z}$, $\tilde{H}_1$, $\tilde{H}_2$, $\tilde{e}$, $\tilde{\mu}$, $\tilde{\tau}$, $\tilde{\nu}$'s, squarks, antisquarks and gluinos have R = -1. The final state in the decay of an R = -1 particle must have R odd, implying that the lightest supersymmetric particle or LSP is stable.

There are relatively few parameters in the minimal model[2]. One can assume a common soft supersymmetry breaking gaugino mass at the unification scale, say $L \ni M_2 \tilde{W}\tilde{W}$. To avoid phenomenologically unacceptable axions, one must introduce a Higgs mixing mass $L \ni \epsilon\tilde{H}_1\tilde{H}_2$. The ratio of Higgs expectation values, $\tan\beta = v_1/v_2$, $v_1 = <H_1>$, $v_2 = <H_2>$ is also a free parameter but can be chosen to be positive without loss of generality. In this notation, it is H_1 which is responsible for up quark masses so that it will be natural to assume $\tan\beta > 1$. These are the only parameters which determine the mass and composition of the LSP. However, for the relic abundance of LSP's, it is necessary to sepcify the Higgs (scalar) masses which depend on one soft supersymmetry breaking parameter say m_0 (which is equal to the pseudoscalar mass). The sfermion masses (which I will assume to be degenerate) must also be specified. Finally (though it is not a parameter of the supersymmetric model) the value of the top quark mass is necessary.

The only neutral fermions with R = -1 which are expected to reside in the LSP are[3] the wino $\tilde{W}^3$, the partner of the 3rd component of the $SU(2)_L$ guage boson; the bino, $\tilde{B}$, the partner of the $U(1)_Y$ gauge boson; and the two neutral Higgsinos $\tilde{H}_1$, and $\tilde{H}_2$. Gluinos are expected to be heavier, $m_{\tilde{g}} = (\frac{\alpha_3}{\alpha})\sin^2\theta_w M_2$ and do not mix with the other states. The sneutrino[4] was a possibility but has been excluded by direct[5], indirect[6] and accelerator[7] searches.

The combination of neutralinos that make up the LSP can be found by diagonalizaing the mass matrix

$$
(\tilde{W}^3, \tilde{B}, \tilde{H}_1^o, \tilde{H}_2^o)
\begin{pmatrix}
M_2 & 0 & -g_2 v_1/\sqrt{2} & g_2 v_2/\sqrt{2} \\
0 & M_1 & g_1 v_1/\sqrt{2} & -g_1 v_2/\sqrt{2} \\
-g_2 v_1/\sqrt{2} & g_1 v_1/\sqrt{2} & 0 & \epsilon \\
g_2 v_2/\sqrt{2} & -g_1 v_2/\sqrt{2} & \epsilon & 0
\end{pmatrix}
\begin{pmatrix}
\tilde{W}^3 \\
\tilde{B} \\
\tilde{H}_1^o \\
\tilde{H}_1^o
\end{pmatrix}
\tag{1}
$$

where M_1 and M_2 are soft supersymmetry breaking terms giving masses to the U(1) and SU(2) gauginos respectively. In a unified theory $M_1 = M_2$ at the unification scale translates to

$$
M_1 = \frac{5}{3} \frac{\alpha_1}{\alpha_2} M_2
\tag{2}
$$

at low energies. By performing a change of basis to $(\tilde{W}^3, \tilde{B}, \tilde{A}^o, \tilde{S}^o)$ with

$$
\tilde{A}^o = \frac{v_1 \tilde{H}_1^o - v_2 \tilde{H}_2^o}{v} = \sin\beta \tilde{H}_1 - \cos\beta \tilde{H}_2
\tag{3}
$$

$$
\tilde{S}^o = \frac{v_2 \tilde{H}_1^o + v_1 \tilde{H}_2^o}{v} = \cos\beta \tilde{H}_1 + \sin\beta \tilde{H}_2
\tag{4}
$$

and $v^2 = v_1^2 + v_2^2$ the mass matrix simplifies and becomes

$$
(\tilde{W}^3, \tilde{B}, \tilde{A}^o, \tilde{S}^o)
\begin{pmatrix}
M_2 & 0 & -g v_2/\sqrt{2} & 0 \\
0 & M_1 & g_1 v_1/\sqrt{2} & 0 \\
-g_2 v/\sqrt{2} & g_1 v/\sqrt{2} & \frac{-2v_1 v_2}{v^2}\epsilon & \frac{v_1^2 - v_2^2}{v^2}\epsilon \\
0 & 0 & \frac{v_1^2 - v_2^2}{v^2}\epsilon & \frac{2v_1 v_2}{v^2}\epsilon
\end{pmatrix}
\begin{pmatrix}
\tilde{W}^3 \\
\tilde{B} \\
\tilde{A}^o \\
\tilde{S}^o
\end{pmatrix}
\tag{5}
$$

and can be solved analytically.

The eigenstates of (1) can be written as

$$\chi_i = \alpha_i \tilde{W}^3 + \beta_i \tilde{B} + \gamma_i \tilde{H}_1^o + \delta_i \tilde{H}_2^o \tag{6}$$

The lightest eigenstate is the LSP. There are some limiting cases in which the LSP is nearly a pure state. When $\epsilon \rightarrow 0$, $\tilde{S}^o$ is the LSP with

$$m_{\tilde{S}} \rightarrow \frac{2v_1 v_2}{v^2} \epsilon \tag{7}$$

When $M_2 \rightarrow 0$, the photino is the LSP[8] defined by

$$\tilde{\gamma} = \frac{g_1 \tilde{W}^3 + g_2 \tilde{B}}{\sqrt{g_1^2 + g_2^2}} = \sin\theta_w \tilde{W}^3 + \cos\theta_w \tilde{B} \tag{8}$$

and

$$m_{\tilde{\gamma}} \rightarrow \frac{8}{3} \frac{g_1^2}{g_1^2 + g_2^2} M_2 \tag{9}$$

When M_2 is large and $\epsilon \gg M_2$, the bino, $\tilde{B}$, is the LSP[9] and

$$m_{\tilde{B}} \simeq M_1 \tag{10}$$

and finally when ϵ is large and $M_2 \gg \epsilon$ either the Higgsino state[9]

$$\tilde{H}_{(12)} = \frac{1}{\sqrt{2}} (\tilde{H}_1^o + \tilde{H}_2^o) \qquad \epsilon > 0 \tag{11}$$

or the state

$$\tilde{H}_{[12]} = \frac{1}{\sqrt{2}} (\tilde{H}_1^o - \tilde{H}_2^o) \qquad \epsilon < 0 \tag{12}$$

is the LSP depending on the sign of ϵ.

The relic abundance of LSP's is determined by solving the Boltzmann equation for the LSP number density in an expanding universe. The technique used[10] is similar to that for computing the relic abundance of massive neutrinos[11]. The ratio of the LSP mass density ρ_χ to the
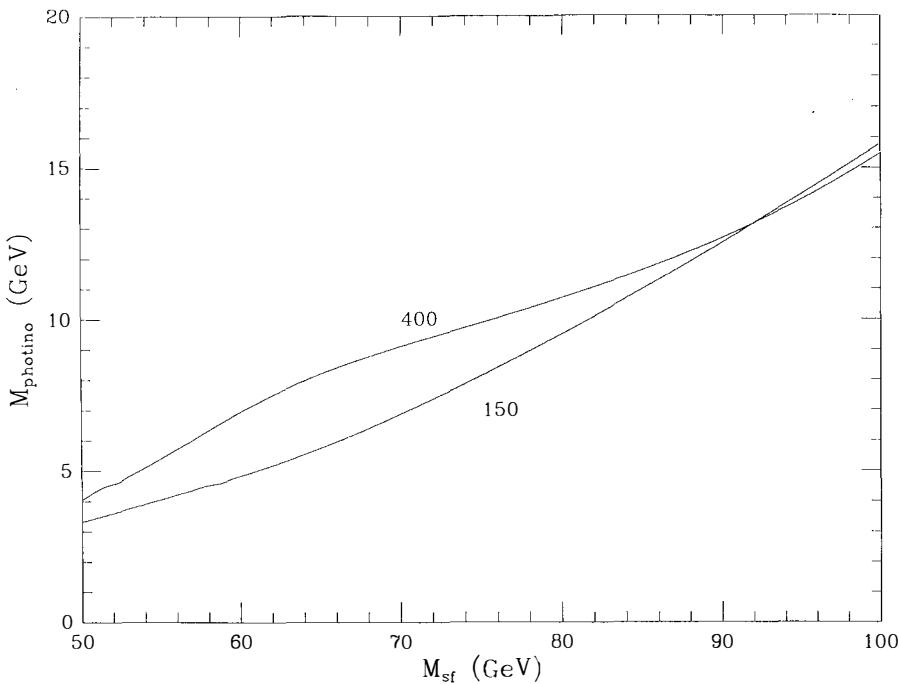
Figure 1

critical density to close the Universe, $\rho_c \approx 1.88 \times 10^{-29}$ h^2 gcm^{-3} ($h \approx$ H/100km Mpc^{-1} s^{-1} is the Hubble parameter), is $\Omega_\chi = \rho_\chi/\rho_c$ and is determined largely by the annihilation cross-section of χ's. The LSP's are in thermal equilibrium until their annihilation rate $\Gamma \propto \langle\sigma v\rangle_A$ falls below the expansion rate of the universe, H. This occurs when $T \sim m_\chi/20$. For example, in the case that the LSP is a photino[8,3,10], the annihilation cross-section depends on the unknown sfermion masses, $\langle\sigma v\rangle_A \propto M_{\tilde{f}}^{-4}$. Thus it is possible to adjust $M_{\tilde{f}}$ so as to insure that $\Omega_\chi h^2 = 1/4$, (corresponding to $\Omega = 1$ and $h = 1/2$). This is shown in Figure 1[10], for two possible quark-hadron transition temperatures, which control the evolution of the number of degrees of freedom and hence the expansion rate. The relic abundance for all LSP's is determined in a similar way, however the annihilation cross-sections may not all have adjustable parameters to insure $\Omega h^2 = 1/4$.

In what follows, I will describe the "pure" dark matter candidates in the standard supersymmetric model, by varying the two mass parameters, M_2
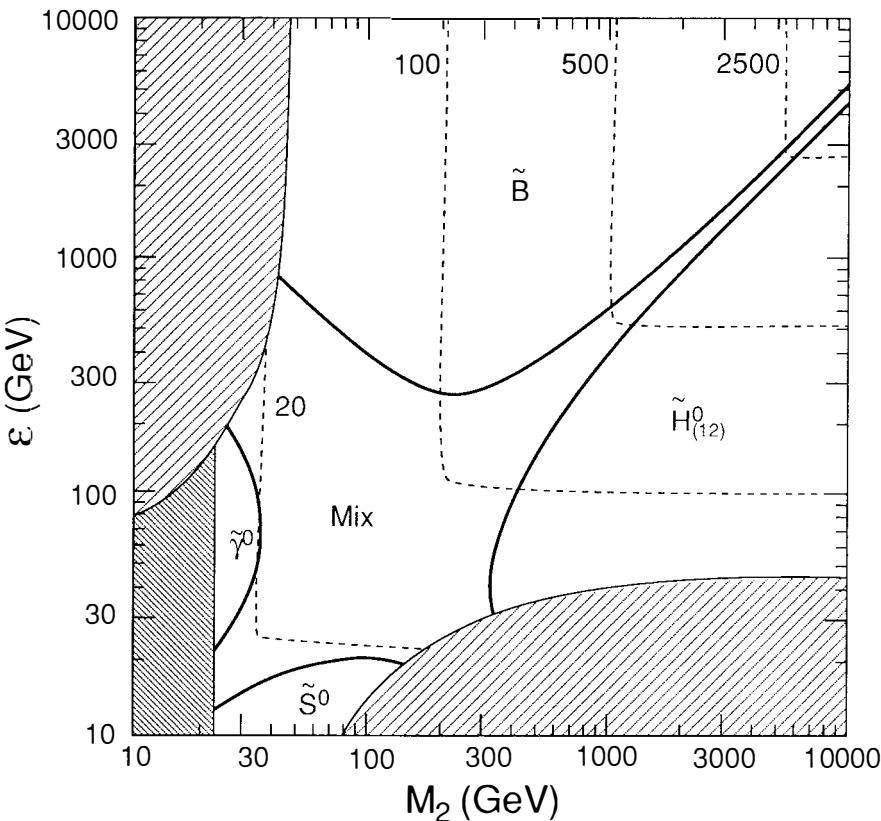
Figure 2

and ϵ from $10\text{-}10^4$ GeV and choosing typical values of $\tan\beta$. In Figure 2[12], I show the regions where the LSP is at least 99% pure along with their mass contours. The light shaded regions correspond to parameters in which there is a chargino ($\widetilde{W}^{\pm}$, $\widetilde{H}^{\pm}$) state with mass $m_{\widetilde{c}} \lesssim 45$ GeV and as such excluded by LEP[13]. The dark shaded region corresponds to a limit on M_2 from the limit[14] on the gluino mass $M_2 \geq 22$ GeV. Notice that the parameter space is dominated by the $\widetilde{B}$ or $\widetilde{H}_{12}$ pure states, and that the photino (most often quoted as the LSP) only occupies a small fraction of the parameter space as does the Higgsino combination $\widetilde{S}^o$ (see Ref. 15 for further constraints on $\widetilde{S}^o$ from LEP). In Figure 2, $\tan\beta = 2$ and $\epsilon > 0$. Figures 3 and 4 show the LSP pure states and masses for $\tan\beta = 2$ with $\epsilon < 0$ and for $\tan\beta = 8$ and $\epsilon > 0$ ($\tan\beta = 8$ with $\epsilon < 0$ is very similar to Figure 4).

The figure shows a plot with axes $-\epsilon$ (GeV) on the vertical axis (ranging from 10 to 10000) and M_2 (GeV) on the horizontal axis (ranging from 10 to 10000). Contour labels include 100, 500, 2500, with regions labeled $\tilde{B}$, Mix, and $\tilde{H}^0_{[12]}$.
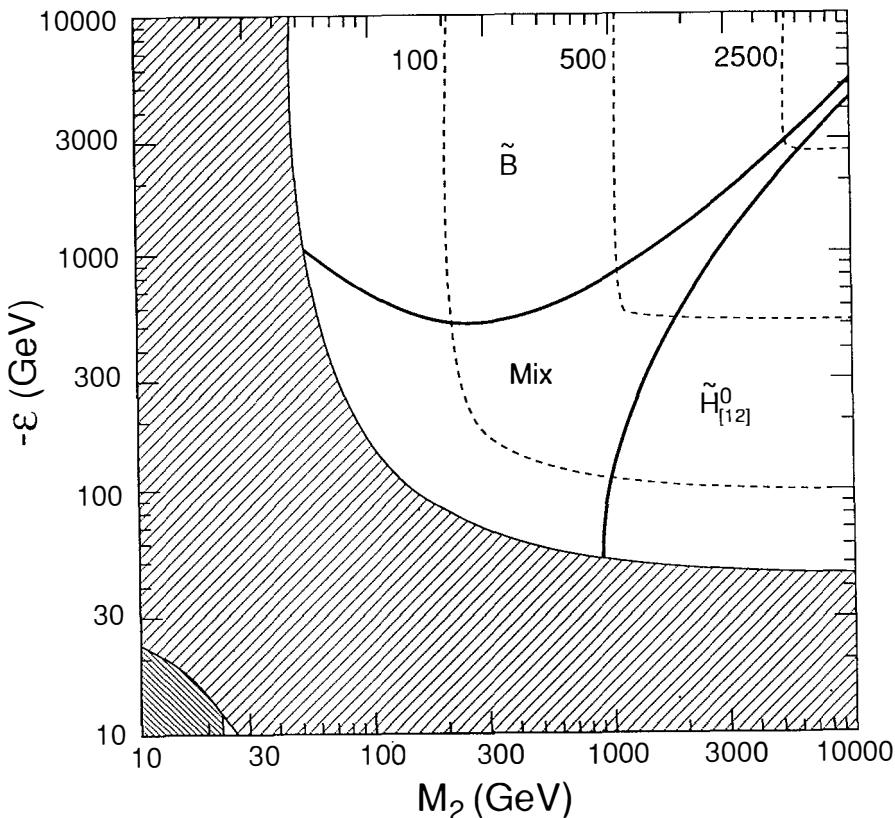
Figure 3

For binos, as was the case for photinos, it is possible to adjust the sfermion mass $M_{\tilde{f}}$ to obtain $\Omega_{\tilde{B}}h^2 = 1/4$. In Figure 5[12], I show the bino region with sfermion mass contours required to make $\Omega_{\tilde{B}}h^2 = 1/4$, for $\tan\beta = 2$, $\epsilon > 0$, the mass of the Higgs pseudoscalar is $m_o = 37$ GeV and the top quark mass is taken to be $m_t = 78$ GeV. The right portion of the bino, parameter space is excluded[6,12,16] because in it, unless $M_{\tilde{f}} < M_{\tilde{B}}$ (contrary to the assumption that the bino is the LSP) $\Omega h^2 > 1/4$. Increasing the top quark mass shifts the $M_{\tilde{f}}$ contours somewhat.

If we fix the sfermion masses, higgs masses and top quark mass we can plot contours of Ωh^2 as is shown in Figure 6, for $\tilde{B}$ and $\tilde{H}_{12}$. In Figure 6, $\tan\beta = 2$, $\epsilon > 0$, $m_o = 37$ GeV, $m_t = 78$ GeV and $M_{\tilde{f}} = 74$ GeV. For binos, one sees the same constraint as in Figure 5. For Higgsinos, much of the lower
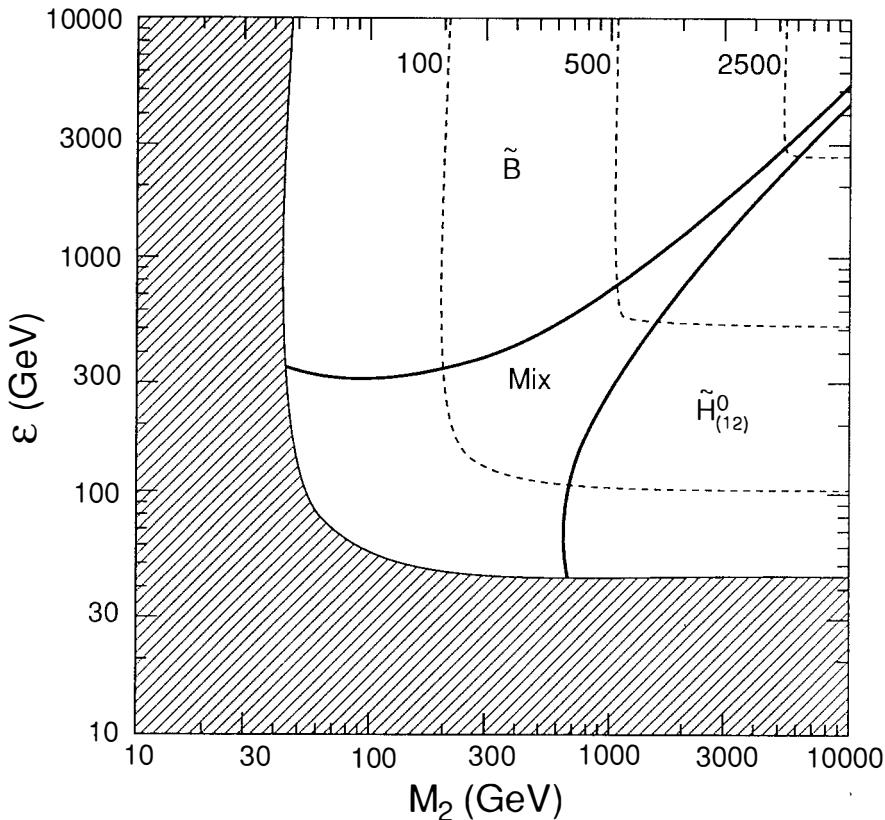
Figure 4

part is excluded as the cross-section there is too small leaving too large a $\tilde{H}_{12}$ abundance. The annihilations there are dominated by Higgs scalar final states and the $\langle\sigma v\rangle_A \propto \dfrac{1}{M_2^2}$, so that for large M_2, annihilations are too slow. The relic abundance of $\tilde{H}_{12}$ changes dramatically as one approaches the $W^{\pm}$ threshold[6,12,16]. Not until $\epsilon \sim 3M_w$ does $\Omega_{\tilde{H}_{12}} h^2$ begin to grow appreciably. For $\epsilon \geq 1$ TeV, the higgsino abundance is again too large.

For $\epsilon < 0$ (holding other parameters fixed) or for $\tan\beta > 2$, the lower excluded region for $\tilde{H}_{12}$ shrinks somewhat. If the top mass is increased, the bino abundance is increased below m_t, whereas for higgsinos the higher top mass keeps $\langle\sigma v\rangle_A$ large at higher ϵ, widening the area with no DM. For
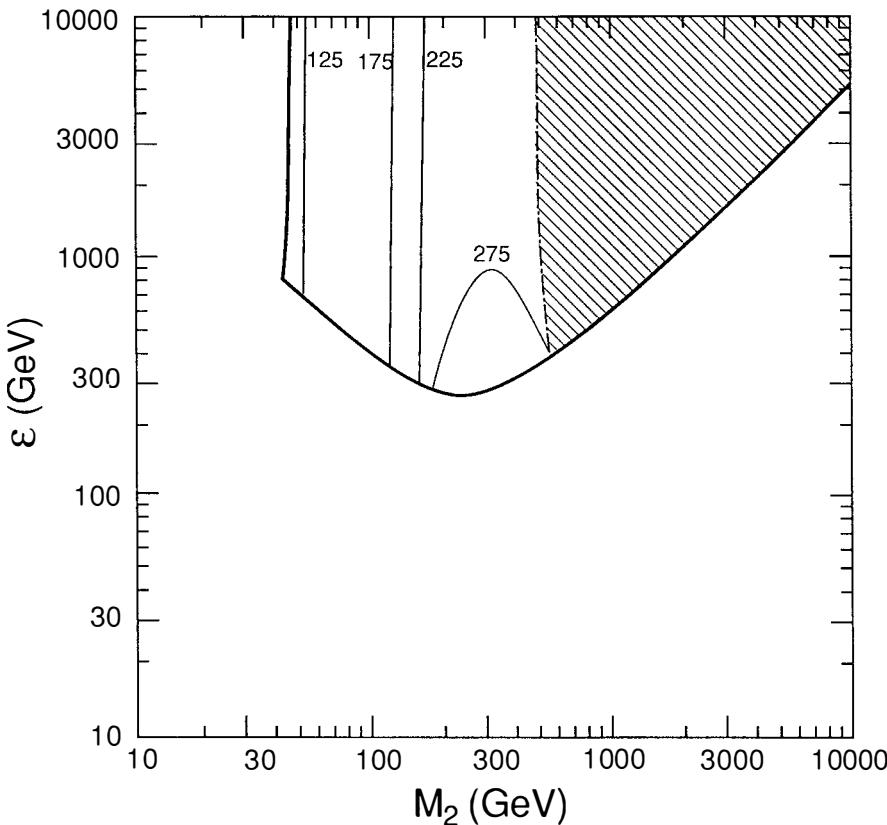
Figure 5

larger higgs masses, the excluded area at the bottom of the $\bar{H}_{12}$ area is enlarged.

The possibilities for cosmological dark matter from LSP's is as follows: For photino, there remains only a small corner of parameter space remaining. First from LEP[17], $\tan\beta \gtrsim 1.2$. For higher values of $\tan\beta$ we find[12)

$$14 \text{ GeV} \lesssim m_{\tilde\gamma} \lesssim 25 \text{ GeV} \qquad \tan\beta = 1.2 \quad \text{any } \epsilon \qquad (13a)$$

$$14 \text{ GeV} \lesssim m_{\tilde\gamma} \lesssim 20 \text{ GeV} \qquad \tan\beta = 2 \quad \epsilon > 0 \qquad (13a)$$

$$\text{none} \qquad \tan\beta \gtrsim 3 \quad \text{any } \epsilon \qquad (13c)$$
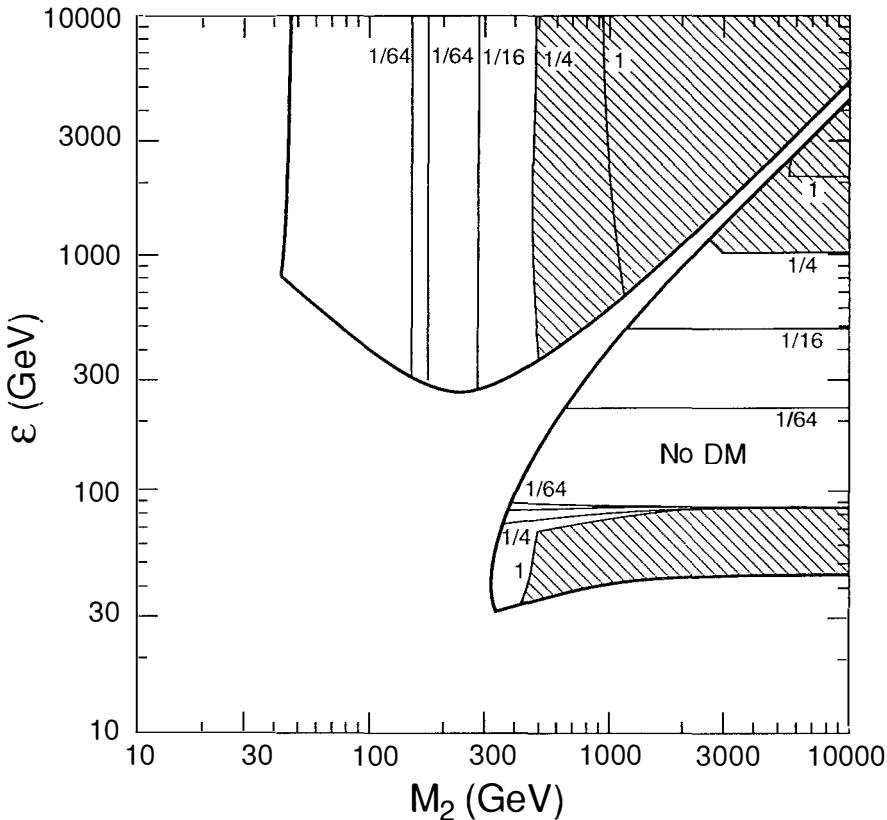
Figure 6

I stress that in the minimal supersymmetric model, there is no such thing (nor ever was) as a 40 GeV let alone a 100 GeV mass photino. For Higgsinos, $\tilde{S}^o$, there is very little possibility left.

Binos offer the best chance for a dark matter candidate in the minimal supersymmetric model. For fixed $M_{\tilde{f}}$, $\Omega_{\tilde{B}} \geq 0.05$ for all $m_{\tilde{B}} = 20\text{-}250$ GeV, and is largely independent of $\tan\beta$ and the sign of ϵ. $m_{\tilde{B}} \gtrsim 250$ GeV is excluded for $\Omega h^2 \lesssim 1/4$ since $M_{\tilde{f}} < m_{\tilde{B}}$ in that case. For fixed $\Omega h^2 = 1/4$, $M_{\tilde{f}} = 120\text{-}300$ GeV for $M_{\tilde{B}}$ in the range 20-250 GeV. The important point is that one can find $\Omega = 1$ for a large portion of the bino parameter space.

For the Higgsinos $\tilde{H}_{12}$, the results are largely independent of $M_{\tilde{f}}$. When ϵ is large $\tilde{H}_{12}\tilde{H}_{12} \rightarrow W^+W^-$, Z^oZ^o is dominant and when ϵ is small $\tilde{H}_{12}\tilde{H}_{12} \rightarrow H^oH^o$ is dominant. Higgsinos in the mass range $m_{\tilde{H}_{12}} \simeq 45 - M_W$ are

allowed only for M_2 less than some critical value of M_2 which is a function of $\tan\beta$. For $m_{\tilde{H}_{12}} \simeq M_W$ - 250 GeV, $\Omega_{\tilde{H}_{12}} \leq 0.05$, and $m_{\tilde{H}_{12}} \simeq 250$-1000 GeV $\Omega_{\tilde{H}_{12}} = 0.05$ - 1.0 (h=1/2) but is relatively unadjustable as it is largely independent of $\tan\beta$. $m_{\tilde{H}_{12}} \gtrsim 1$ TeV is excluded.

Finally, I would like to comment on the detectability of $\tilde{H}_{12}$'s and $\tilde{B}$'s[18]. Binos interact in much the same way as do photinos. The photino elastic cross section for $\tilde{\gamma} + p \rightarrow \tilde{\gamma} + p$ is

$$\sigma_{\tilde{\gamma}} = \frac{3}{\pi} \frac{1}{M_{\tilde{f}}^4} f_{\tilde{\gamma}}^2 \frac{m_{\tilde{\gamma}}^2 m_p^2}{(m_{\tilde{\gamma}}+m_p)^2}$$

$$f_{\tilde{\gamma}} = 4\pi \alpha \Sigma Q_q^2 \Delta q$$

where Δq is related to the spin of the proton ($S_p^\mu \Delta q = <p,s|S_p^\mu|p,s>$, $S_p^\mu = \frac{1}{2} \bar{q} \gamma^\mu \gamma_5 q$). For binos, $m_{\tilde{\gamma}} \rightarrow m_{\tilde{B}}$ and $f_{\tilde{\gamma}} \rightarrow f_{\tilde{B}}$ with

$$f_{\tilde{B}} = 4\pi\alpha_1 \Sigma \frac{Y_{q_L}^2 + Y_{q_R}^2}{8} \Delta q$$

where α_1 is the U(1) hypercharge fine structure construct.. In the naive quark model, $\sigma_{\tilde{B}}/\sigma_{\tilde{\gamma}} = 0.46$.

For Higgsinos, we can look at the elastic cross-section for $\tilde{S}$,

$$\sigma_{\tilde{S}} = \frac{3}{2\pi} G_F^2 \cos^2 2\beta \ g_A^2 \frac{m_{\tilde{S}}^2 m_p^2}{(m_{\tilde{S}}+m_p)^2}$$

(when $\cos 2\beta = 1$, this is the Majorana neutrino scattering cross-section). However for $\tilde{H}_{12}$, $\cos 2\beta = 0$! and the dominant effects come from impurities in $\tilde{H}_{12}$, although for $\geq 99\%$ pure $\tilde{H}_{12}$, $\sigma_{\tilde{H}_{12}}/\sigma_{\tilde{\gamma}_m} \lesssim 0.08$, making detection of these candidates rather difficult.

REFERENCES

1. see: J.R. Primack, _Enrico Fermi, Course 92_, N. Cabibbo, ed. (North Holland, Amsterdam, 1987) p. 137; V. Trimble, Ann. Rev. Astron.

Astrophys. 25 425 (1987); Dark Matter, M. Srednicki, ed. (North Holland, Amsterdam, 1989) p. ; J. Primack, D. Seckel and B. Sadulet, Ann. Rev. Nucl. Part. Sci. 38, 751 (1988).

2. H.E. Haber and G.L. Kane, Phys. Rep. 117, 75 (1985); J.F. Gunion and H.E. Haber, Nucl. Phys. B272, 1 (1986).

3. J. Ellis, J. Hagelin, D.V. Nanopoulos, K.A. Olive and M. Srednicki, Nucl. Phys. B238, 453 (1984).

4. L.E. Ibanez, Phys. Lett. 137B, 160 (1984); J. Hagelin, G.L. Kane and S. Raby, Nucl. Phys. B241, 638 (1984).

5. S. Ahlen et al., Phys. Lett. B195, 603 (1987); D.D. Caldwell et al., Phys. Rev. Lett. 61, 510 (1988).

6. K.A. Olive and M. Srednicki, Phys. Lett. 205B, 553 (1988).

7. see e.g. B. Adeva et al., Phys. Lett. B231, 509 (1989); D. Decamp et al., Phys. Lett. B231, 519 (1989); M.Z. Akrawy et al., Phys. Lett. B231, 530 (1989); P. Aernio, Phys. Lett. B231, 539 (1989).

8. H. Golberg, Phys. Rev. Lett. 50, 1419 (1983); L.M. Krauss, Nucl. Phys. B227, 556 (1983).

9. K.A. Olive and M. Srednicki, Phys. Lett. B230, 78 (1989).

10. M. Srednicki, R. Watkins and K.A. Olive, Nucl. Phys. B310, 693 (1988).

11. B. Lee and S. Weinberg, Phys. Rev. Lett. 39, 165 (1977).

12. K.A. Olive and M. Srednicki, University of Minnesota preprint UMN-TH-805/90 (1990).

13. M.Z. Akrawy et al., Phys. Lett. B (in press) 1990; D. Decamp et al., Phys. Lett. B (in press) 1990.

14. J. Alitti et al., Phys. Lett B235, 363 (1990).

15. J. Ellis, G. Ridolfi and F. Zwirner, CERN preprint, TH.5596/89 (1989); L.M. Krauss, Phys. Rev. Lett. 64, 999 (1990) and these proceedings.

16. K. Greist, M. Kamionowski and M.S. Turner, Fermilab preprint 89-239 (1989).

17. D. Decamp et al., Phys. Lett. B (in press) 1990.

18. R. Flores, K.A. Olive and M. Srednicki, Phys. Lett. B237, 72 (1990).

NEUTRINO MAGNETIC MOMENT

Darwin Chang[a,b] and Goran Senjanović[c]

a) Dept. of Physics and Astronomy, Northwestern
   Univ., Evanston, IL 60208
b) Fermi National Laboratory, Batavia, IL 60510
c) Theoretical Physics Dept. University of
   Zagreb, 41000 Zagreb

ABSTRACT

We review attempts to achieve a large neutrino magnetic moment ($\mu_\nu \gtrsim 10^{-11} \mu_B$), while keeping neutrino light or massless. The application to the solar neutrino puzzle is discussed.

INTRODUCTION

The so called solar neutrino puzzle is now more than twenty years old. The experimenrally detected[1] flux of neutrinos originating from the thermonuclear reactions in the sun via the process

$$\nu_e^{37}Cl \rightarrow e^{-37}A_r \tag{1}$$
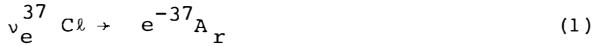
is far below the theoretical predictions.[2] Whereas the standard Solar Model (SSM) suggests the conversion rate

$$<{}^{37}Cl>_{SSM} = (7.9 \pm 2.6) \; SNU \tag{2}$$

the experiments give a rate roughly a factor of three smaller

$$<{}^{37}Cl>_{exp} = (2.1 \pm 0.3) \; SNU \tag{3}$$

The unit 1 SNU (Solar Neutrino Unit) corresponds to 10^{-36} captures/atom-sec. For the experiment of Davis et al 1 SNU is equivalent to the conversion of about 0.23 atoms of $^{37}A_r$ per day, so that (3) gives a signal of about 0.5 atoms per day. This is a rather challenging experiment and we won´t go into the uncertainties associated with it or the theory, we suggest Peccei´s review for a nice discussion of whether this is a true paradox or not. The beauty of this issue is it touches upon the fundamental properties of neutrinos such as their masses, mixing angles, magnetic moments etc.

The most popular explanation to date is the MSW mechanism[4] of resonant oscillations in the interior of the sun, which transform ν_e into ν_μ (or ν_τ). For this to work, the masses and the mixing angles are constrained by

$$\Delta m^2 = m_{\nu_\mu}^2 \; (m_{\nu_\tau}^2) - m_{\nu_e}^2 \sim 10^{-7} - 10^{-4} (eV)^2$$
$$\sin^2 \theta \gtrsim 10^{-3} \tag{4}$$

We should stress that the MSW process takes place in the radiation zone (0.04 - 0.7 $R_\odot$, where $R_\odot$ is the radius of the sun). This perfectly acceptable explanation needs

further laboratory verification, for we still have no clue on the values of neutrino masses and mixing angles. A potential death blow to this idea could come, however, from the very observations of Davis et al. They seem to suggest an anticorrelation between solar neutrino flux and the sun-spot activity. At the sunspot maximum, the neutrino flux is below average and at sun-spot minimum activity it is far above the average value: (5.1 ±1) SNU (1987-1988). The sun-spot activity is attributed to magnetic storms in the convection zone of the sun, near the surface (0.7 - 1 $R_{\odot}$), when the magnetic field near the maximum reaches the value of $10^3 - 10^4$ Gauss. If this is so, then the mass oscillation cannot be the explanation; it has to do with the magnetic field of the sun.

Cisneros[5] and Voloshin[6] et al (VVO) suggests that the electron neutrino possesses a large magnetic moment which enables it to flip into the sterile right - handed component (or other flavors) in the magnetic field of the sun. This seems to work if the magnetic moment is on the order of $10^{-11} \mu_B$ ($\mu_B = \frac{e}{2m_e}$), which, as we will see below, is an enormous number, many orders of magnitude above the value in the standard model (with ν_R).

NEUTRINO MAGNETIC MOMENT

Recall that, like the mass, magnetic moment is the helicity flip operator

$$\mu_\nu \bar{\nu}_R \sigma^{\mu\nu} \nu_L F_{\mu\nu} = \mu_\nu \nu_L^T C \sigma^{\mu\nu} \nu_L^C F_{\mu\nu} \tag{5}$$

For the VVO mechanism to work, μ_ν is estimated[6]

$$\mu_\nu \sim 10^{-11} - 10^{-10} \mu_B \tag{6}$$

To get a feeling for this number, let us campare it with the natural value for μ_ν in the standard model[7] (with ν_R)

$$(\mu_\nu)_{st} = \frac{3eG_F m_\nu}{8\sqrt{2}\ \pi^2} = 3.2 \times 10^{-19}\ (\frac{m_\nu}{1ev})\mu_B \tag{7}$$

which is at least eight orders of magnitude smaller than the VVO number!   This number can get substantially enhanced if there is a right-handed current, such as in left-right models[8]

$$(\mu_\nu)_{LR} = \frac{G_F\ m_e}{2\sqrt{2}\ \pi^2}\ \sin 2\Phi \underset{\sim}{} 2 \times 10^{-13}\ \sin 2\Phi\ \mu_B \tag{8}$$

where Φ is the mixing angle between left and right-handed currents, $|\Phi| \lesssim 0.05$. Obviously, (8) is still too small. With some optimal assumptions supersymmetry may give even larger prediction, but still

$$(\mu_\nu)_{SS} \lesssim 10^{-12}\ \mu_B \tag{9}$$

Thus we have problem at our hands: how to reconcile a large value for μ_ν with a small neutrino mass, $m_\nu \lesssim 10eV$? This is quite a challenge for a model builder, one which is hard to ignore, even if you are not very convinced with the VVO mechanism. The idea, of course, is to uncorrelate μ_ν with m_ν, which was achieved by Babu and Mathur and Fukugita and Yanagida.[9] They introduce an SU(2) singlet, charged Higgs field h^+, which works OK for μ_ν . The trouble is that in their models neutrino has a bare Dirac mass, which has to be fine tuned to be small.[10] This is not a solution to our problem.

An ingenious way out was suggested by Voloshin[11], who postulated an $SU(2)_H$ symmetry between ν and ν^C. Under $SU(2)_H$ the mass term behaves as a triplet and so is forbidden, whereas the magnetic moment interaction is allowed. Namely,

$$L_\nu^T\ i\ \sigma_2\ CL_\nu = 0$$
$$L_\nu^T\ i\ \sigma_2\ C\ \sigma_{\mu\nu}L_\nu \neq 0 \tag{10}$$

where $L_\nu = (\nu_{\nu^C})_L$. This idea, unfortunately is not easy to realize, since $SU(2)_H$ does not commute with $SU(2)_L$. You

could try to enlarge $SU(2)_L \times U(1)$ a la Barbieri and Mohapatra[12], who use $SU(3)_C \times U(1)$ with the basic triplet consisting of e, ν and ν^C. This automatically contains $SU(2)_H$, but the scale of $SU(2)_H$ symmetry breaking is pushed above M_W! This means that $SU(2)_H$ effectively loses its original role; we are back where we started.

This may be a good place to mention some potential troubles with the large value of μ_ν.

(a) laboratory experimental[13] limit on μ_ν is

$$(\mu_\nu)_{exp} \leq 1.5 \times 10^{-10} \ \mu_B \tag{11}$$

(b) from the stellar cooling analysis, there is an astrophysics bound[14]

$$(\mu_\nu)_{astro} \lesssim 1.1 \times 10^{-11} \ \mu_B \tag{12}$$

(c) also, cosmological argument due to increase in ν_L number from the $\nu_R + e \to \nu_L + e$ process, induced via μ_ν, implies[14]

$$(\mu_\nu)_{cosmol} \lesssim 0.5 \times 10^{-11} \ \mu_B \tag{13}$$

(d) and finally, the energy loss of the supernovae 1987a would be too large, unless[15]

$$(\mu_\nu)_{sn} \lesssim 10^{-12} \ \mu_B \tag{14}$$

Obviously the serious problem is d) only, but there one could try to add new interactions which would trap ν_R, so even this is not a fatal blow to the VVO suggestion. Still, we find the idea of transition magnetic moments far more appealing. Here, in analogy with neutrino oscillations, one takes $\nu_{\mu L}$ instead of ν_{eL}^C. In other words, one assumes the horizontal symmetry[16] $SU(2)_H$ between the Weyl fields ν_e and ν_μ, so that now

$$L_\nu = \begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix}_L \tag{15}$$

The invariant magnetic moment (10), with L_ν from (15) is now a transition moment

$$\nu_e^T C \sigma_{\mu\nu} \nu_\mu F^{\mu\nu} \neq 0 \tag{16}$$

The diagonal moments vanish in the case of Weyl (Majorana) neutrinos.

The new problem appears, though; the magnetic field of the sun has no energy to flip the neutrinos, unless the mass difference between ν_e and ν_μ is almost vanishing[17]

$$\Delta m^2 = m^2_{\nu_\mu} - m^2_{\nu_e} \lesssim 10^{-7} (eV)^2 \tag{17}$$

The model building becomes even more challenging. The most natural approach is to use some symmetry, such as U(1) global symmetry $L_e - L_\mu$, which would imply $\Delta m^2 = 0$. In this case, ν_e and ν_μ group together into a Dirac neutrino. However, we still have to achieve $m_\nu \lesssim 10 eV$ or so.

Now $SU(2)_H$ could be either global or local, however, in both cases its scale of symmetry breaking M_H must be large. If it is global, the consequence of Goldstone bosons implies $M_H \geq 10^7 - 10^8$ GeV, and the local symmetry case demands $M_H > M_W$. But this defies the whole purpose for imposing $SU(2)_H$ in the first place; we need this symmetry at low energies, so that it can play its custodial role. Ideally, $M_H << 1$. GeV is what we are after. We are back where we started. What to do?

A suggestion by Leurer and Marcus[17] is to have an approximate $SU(2)_H$, broken explicitly. Encouraged by the fact that in the standard model $SU(2)_H$ is broken only by

$$\epsilon = \frac{m_\mu - m_e}{M_H} \sim 10^{-4} \text{ terms, they demand the breaking of}$$

this symmetry to continue being proportional to ϵ , a rather ad hoc assumption. This would enable them to lower the value of

m_ν down to eV scale.

Leurer and Golden[19] try a U(1), instead of SU(2)$_H$ symmetry. This also has its problems, in particular they need some further fine tuning to achieve a small mass for m_ν.

Yet another possibility is to use a discrete symmetry. The discrete symmetry certainly needs a serious motivation, for it carries a potential disaster of domain walls. To get rid of this, we appeal to the idea of possible symmetry non-restoration at high temperature, as advocated by Weinberg and especially Mohapatra and G.S.[20] (or soft breaking, if you find the above hard to digest). The main advantage of the discrete symmetry is that it can be kept down to as low energies as we wish (maybe even left unbroken?) Besides us, the discrete symmetry approach was also tried by Babu and Mohapatra[21] and by the Wienna group[22]. Babu and Mohapatra argue that the combination of supersymmetry and the Z_4 subgroup of $L_e - L_\mu$ works, whereas in the case of Wienna group Dirac neutrino ends up light for somewhat accidental resons (but technically natural). The space doesn't permit as to discuss their work at great length; for reasons that you will probably find obvious we shall devote the rest of this paper to our work.

THE DISCRETE SYMMETRY AND μ_ν

The first fact to notice is that the custodial symmetry must be nonabelian, if it is to forbid $\nu_i^T \, C \, \nu_j$ and allow $\nu_i^T \, C \, \sigma_{\mu\nu} \, \nu_j$. The simplest possibility would be a subgroup of Voloshin's SU(2)$_H$ symmetry which does the job. The ideal candidate for this is the quaternion symmetry Q_4 of 8 elements, whose characteristics are best read off from its faithfull two dimensional representation

$$\{ \pm 1_2, \ \pm i \ \sigma_i \}$$

(a) Quaternion symmetry Q_4

Since there is always a trivial repr. (R_1), obviously

there are four one dimensional representations $R_1, \ldots R_4$. The five classes ($C_1 = \{1_1\}$, $C_2 = \{-1_2\}$, $C_3 = \{\pm i\sigma_1\}$, $C_4 = \{\pm i\sigma_2\}$, $C_5 = \{\pm i\sigma_3\}$ in the case of a doublet representation D) are then given by the character table

	C_1	C_2	C_3	C_4	C_5
R_1	1	1	1	1	1
R_2	1	1	1	-1	-1
R_3	1	1	-1	1	-1
R_4	1	1	-1	-1	1
$R_5 = D$	2	-2	0	0	0

Following Voloshin, we assume $L = \binom{\nu_e}{\nu_\mu}$ to be in a doublet representation of Q_4. In the same manner as with $SU(2)_H$ symmetry (see eq. (10)), the mass term is forbidden, whereas magnetic moment is allowed. The point is that the antisymmetric transformation $D^T i\sigma_2 D$ is invariant, which vanishes for m_ν. One can now construct one's favourite model; all you need the decomposition

$$(D \times D)_{AS} = R_1$$

$$(D \times D)_S = R_2 + R_3 + R_4$$

Obviously, both new fermions and Higgs fields are needed, and it is somewhat a matter of taste of what to choose.

We shall not describe the details of our model here[23], for they are given in our paper. Let us rather concentrate on the shortcomings of the Q_4 approach. The main problem is the fine tuning needed to keep Δm_ν^2 small. The way we construct the model, there is no protective symmetry which could make ν_e, ν_μ a Dirac neutrino. The potential candidate is a Z_4 symmetry generated by, say $i\sigma_1$ (or any other $i\sigma_k$), which should remain unbroken. Therefore, only those Higgs fields transforming as R_1 or R_2 are allowed nonvanishing ver's. This forces e_R, μ_R to form a doublet D and the outcome is that we cannot split e, μ masses. This problem seems to remain even if we add additional particles.

(b) dicyclic group Q_6

This group consisting of 12 elements is more promising,

since it has two doublet representations. It is generated by elements r,s through

$$r^6 = s^4 = 1, \quad srs = r^2 \tag{18}$$

Its ´six classes $C_1\{1\}$, $C_2\{-1\}$, $C_3\{r^2,-r\}$, $C_4\{s,-sr,sr^2\}$, $C_5\{-s,sr,-sr^2\}$,$C_6\{r,-r\}$ has the following character table

	C_1	C_2	C_3	C_4	C_5	C_6
R_1	1	1	1	1	1	1
R_2	1	1	1	-1	-1	1
R_3	1	-1	1	i	-i	-1
R_4	1	-1	1	-i	i	-1
R_5	2	-2	-1	0	0	1
R_6	2	2	1	0	0	-1

The explicit form of r and s for R_5,R_6 is

$$r_5 = \exp\left(\frac{i\, 2\pi\sigma_1}{6}\right), \quad s_5 = i\sigma_3$$

$$r_6 = \exp\left(\frac{i\, 2\pi\sigma_1}{3}\right), \quad s_6 = -\sigma_3 \tag{19}$$

It is clear that Q_6 is not a subgroup of $SU(2)_H$, but rather $U(2)_H$. However, it still forbids the neutrino mass, if we assign left-handed leptons into R_5 (or $(R_5)^{\dagger}$).

Our strategy is straightforward: we break Q_6 down to Z_4 generated by s. This implies neutrino mass term $\nu_e^T C \nu_\mu$, or $\Delta m_\nu^2 = 0$. Furthermore, if you choose

$$e_R(R_4), \quad \mu_R(R_2)$$

$$\psi = \begin{pmatrix} \nu_e & \nu_\mu \\ e & \mu \end{pmatrix} (R_5)^{\dagger} \tag{20}$$

then under s

$$e_{L,R} \to -i\, e_{L,R} \; ; \quad \mu_{L,R} \to i\, \mu_{L,R} \tag{21}$$

In other words, s is a Z_4 subgroup of $L_\mu - L_e$. The dangerous processes such as $\mu \to e\gamma, \mu \to ee\bar{e}$ are forbidden.

Besides the usual particles, this approach needs additional lepton doublets $N_{L,R}$ and new Higgs fields in order to generate large μ_ν. As in our Q_4 work[23], the neutrino (now Dirac) mass is suppressed by m_Q/M_W, where m_Q is the scale of Q_6 symmetry breaking. In order to split $m_\mu - m_e$, $m_Q \sim m_\mu$, giving us m_ν in the eV region.

The technical details of our model will be spelled out elsewhere[24], together with the phenomenological implications. It´s major fault is the proliferation of new fields, but it seems to work. We hope, though, that a more elegant solution along these lines will soon emerge. Needless to say, new experimental results regarding the sun-spot activity and neutrino solar flux are badly needed to tell us whether the anticorrelation between the two is real.

REFERENCES AND FOOTNOTES

1. R.Davis,T.Cleveland and J.K.Rowley, in Steamboat Springs 1984, Proceedings,Interactions Between Particle and Nuclear Physics, pp 1037-1050
2. I.N.Bahcall et al, Rev.Mod.Phys. 54,767(1982); I.N.Bahcall and R.K.Ulrich,Rev.Mod.Phys.60,297(1988)
3. R.D.Peccei, The physics of neutrinos, DESY preprint 89-043 (1989). This nice review of neutrino physics offers also a comprehensive set of references.
4. L.Wolfenstein, Phys.Rev. D17,2369(1978);D20,2634(1979); S.P.Mikheyev and A.Yu.Smirnov, Nuovo Cim. 9C,17(1986)
5. A.Cisneros, Astrophys.Space Sci. 10,87(1981)
6. L.B.Okun,M.B.Voloshin and M.I.Vysotsky,Soc.J.Nucl.Phys.91, 754(1986);Sov.Phys. JETP 64,446(1986)
7. K.Fujikawa and R.E.Shrock,Phys.Rev.Lett.45,963(1980); B.W. Lynn,Phys.Rev. D23,2151 (1981)
8. J.E.Kim, Phys.Rev.D14,3000(1976); M.A.B.Bég,W.J.Marciano and M.Ruderman,Phys.Rev.D17,1395(1978); R.E.Shrock,Nucl.Phys. B206,359(1982); J.Liu,Phys.Rev. D35,3447 (1987) M.J.Duncan et al, Phys.Lett. 191B,304(1987)
9. K.S.Babu and V.Mathur,Phys.Lett. 196B,218(1987);M.Fukugita and T.Yanagida,Phys.Rev.Lett. 58,1807(1987)
10. J.Pulido and J.Ralston,Phys.Rev. D38,2864(1988) J.Liu,Phys.Lett. 225B, 148(1989)
11. M.Voloshin,Sov.J.Nucl.Phys. 48,512 (1988)
12. R.Barbieri and R.N.Mohapatra, Phys.Lett. 218B,225(1989)
13. A.V.Kyuldjiev, Nucl.Phys. B243, 387 (1984)
14. M.Fukugita and S.Yazaki,Phys.Rev. D36,3817(1987)
15. R.Barbieri and R.N.Mohapatra, Phys.Rev.Lett. 61,27(1988) J.Lattimer and J.Cooperstein, ibid 61,23(1988)
16. K.S.Babu and R.N.Mohapatra,Phys.Rev.Lett.63,228(1989)
17. E.Akhmedov,Phys.Lett. 213B,64(1988); C.-S.Lim and W.J.Marciano,Phys.Rev. D37,1368(1988)
18. M.Leurer and N.Marcus,Technion preprint PH-89-48(1989)
19. M.Leurer and M.Golden,Fermilab preprint 89/175-T(1989)
20. S.Weinberg,R.N.Mohapatra and G.Senjanović,Phys.Rev.D20,3390 (1979)
21. K.S.Babu and R.N.Mohapatra, Phys.Rev.Lett.
22. G.Ecker, W.Grimus, H.Neufeld,CERN preprint - TH.5485/89
23. D.Chang,W.-Y.Keung and G.Senjanović,Phys.Rev.D (to be publ.)
24. D.Chang,W.-Y.Keung,S.Lipovača and G.Senjanović,to appear

WIMPS AS A STAR COOLING SYSTEM

François Martin de Volnay

Alain Bouquet

Jean Kaplan.

LPTHE, Université PARIS VI-VII,

2 Place Jussieu, F-75251 Paris CEDEX 05. France.

Abstract

Weakly Interacting Massive Particles are hypothetical particle-candidates for solving the dark matter problem. For a specific range of masses and cross-sections on baryonic matter they are trapped and accumulate in the core of stars. For a smaller range they happen to be very efficient to cool those cores. For the sun, that could be an explanation for the low rate of detected solar neutrinos. If so, constraints on WIMP (cosmion) parameters are very stringent and can be currently tested by ongoing dark matter experiments.

1 Introduction

The acronym WIMP stands for Weakly Interacting Massive Particle. It is a generic word used for any exotic particle solving the dark matter problem. As the word itself means it, WIMP would be a massive particle of a few GeV whose interaction with ordinary matter would be in the realm of weakness. So far no known particle has been discovered to be a WIMP. If in addition WIMPs solve the solar neutrino problem they are called cosmions. All this terminology is matter of convention.

After a preliminary section in which we remind what is the solar neutrino problem, we consider how WIMPs can be captured by stars and concentrated in their core . Then we show how those particles are very efficient to cool the core of stars. In fine we apply this to the sun and its neutrino problem and get strong constraints on the WIMP mass and its cross-section on matter.

2 The Solar Neutrino Problem

Detection of solar neutrinos over 18 years by the Chlorine experiment [1] gives an average of 0.43 neutrinos detected per day, which converted in Solar Neutrino Units gives a rate of 2.3±0.3 SNU. Two of the major standard solar evolution codes give respectively 7.9±0.9 SNU for Bahcall and Ulrich [2] and 5.8±1.2 SNU for the Saclay code [3]. Despite differences between them, those two predictions are roughly a factor 3 above the observed rate. The solar neutrinos detected by the Chlorine experiment come from the reaction $eB^8 \rightarrow \nu_e Be^8$ which takes place in the core of the sun and whose rate varies as T^{21}. On the other hand the main fusion reaction which starts with $pp \rightarrow e^+\nu_e D$ varies as T^5. Therefore a small decrease of the central temperature of the sun, e.g. by a few per cent : 3 to 8 %, will drastically decrease the rate of the Boron reaction and of its corresponding neutrino without changing the pp fusion rate. This does offer a solution to the Solar Neutrino Problem [4,5,6]. If they exist, and with appropriate values of unknown quantities like their mass and cross-section, WIMPs could be very efficient to cool the core of stars (see Section 4) and consequently could give the adequate decrease of the sun central temperature leading to a Boron neutrino rate in accordance with observations.

3 Capture of WIMPs by stars

In order to explain the rotation of spiral galaxies, it is necessary to suppose the existence of a halo of dark matter surrounding those galaxies. For the Milky Way, our own galaxy, if we assume that its halo is an isothermal gas of WIMPs, the distribution of their velocities is gaussian : $\exp[-3v^2/2v_{halo}^2]$, with a mean velocity v_{halo} of the order of 300 km/s (with roughly an error of 20%). In the vicinity of the sun, the WIMP mass density ρ_{halo} is of the order of 0.4 GeV/cm^3, the error being roughly a factor 2.

When a WIMP crosses the surface of a star and gets inside it, depending on its cross-section on matter, the WIMP can eventually interact with a nucleus and looses its kinetic energy. After a few of those interactions and because of the gravitational well of the star the WIMP can be trapped in its core. There is a critical value σ_c of

the WIMP-matter cross-section : for cross-sections larger than this critical value, any WIMP going inside the star will interact and be captured. This critical cross-section σ_c corresponds to a WIMP mean free path of the order of the star radius :

$$\sigma_c \sim m_p R^2 / M$$

where m_p is the proton mass, R the star radius and M its mass. For the sun $\sigma_c = 4 \ 10^{-36} cm^2$. But in order to be really trapped in the core of the sun, WIMPs should not evaporate, i.e. by interaction with matter they should not gain enough kinetic energy allowing them to get out of the star. This question has been extensively studied by Gould [7] who showed that the evaporation rate is exponentially suppressed for WIMP masses larger than an evaporation mass m_{ev} which depends on the WIMP-matter cross-section σ_{xN}[1] : WIMP with mass $m_x \geq m_{ev}(\sigma_{xN})$ are trapped in the core of the star. The evaporation mass m_{ev} is a function of σ_{xN}. In the case of the sun : for $\sigma_{xN} = 4 \ 10^{-36} cm^2$, $m_{ev} \simeq 4 GeV$; whereas for $\sigma_{xN} = 10^{-34} cm^2$, $m_{ev} \simeq 2 GeV$.

If WIMPs are trapped in the core of a star, in order for them to stay there in sufficient number, it is necessary to assume also that they do not annihilate. So photinos, if they do exist, are not good candidates to accumulate in the center of the sun. Non-evaporation and non-annihilation conditions being assumed, the gravitational capture rate of WIMPs by a star is proportional to [6,8,9]

$$\frac{\rho_{halo}}{v_{halo}} \sigma_{xN} \frac{1}{m_x}$$

the proportionality to σ_{xN} occurring only for $\sigma_{xN} \leq \sigma_c$. For a WIMP with mass 4 GeV and cross-section $\sigma_{xN} = 4 \ 10^{-36} cm^2$, the number of WIMPs captured by the sun is of the order of 3 10^{29} per second. After 4.6 Gigayears, the sun lifetime, this leads to approximately 4 10^{46} WIMPs presently trapped inside the sun. The WIMP radial distribution is approximately gaussian being proportional to $\exp[-r^2/r_x^2]$. WIMPs are concentrated in a central region of radius a few r_x which is small as compared to the star radius R. In the case of the sun and for a WIMP mass of 4GeV : $r_x \sim 0.05 \ R$. In this central region the number of WIMPs remains very small compared to the number of hydrogen nuclei (10^9 smaller) and to the number of photons (10^5 smaller), but WIMPs being very efficient to transport energy, much more than photons, despite the smallness of their number, they are able to cool the core of stars.

4   Energy transport by WIMPs

There exist two main regimes of energy transport by WIMPs :

4.1   Non local energy transport or Knudsen regime

In this regime the WIMP mean free path l_x is of the order of (or greater than) the dimension of the region in which they are trapped. The WIMP mean free path is

---

[1]The notation σ_{xN} is correct only if the star is composed of only one type of nucleus N. In realistic cases it should be replaced by an effective WIMP-matter cross-section which depends on the various WIMP-nucleus cross-sections and on the composition of the star [8].

inversely proportional to the WIMP-matter cross-section :

$$l_x(r) = \frac{1}{n_N(r)\ \sigma_{xN}}$$

which is valid for a star made with only one type of nucleus N, whose number density at the distance r from the center is $n_N(r)$ (for a more practical formula taking account of all different species of nuclei composing a star see ref. [10]). The Knudsen regime is therefore defined by the condition $l_x \geq r_x$. It is useful to define a Knudsen number

$$K_n = \frac{l_x(r)}{r_x}$$

The Knudsen regime corresponds to K_n of the order or larger than 1.

In this regime a WIMP orbiting in the central region of extension order r_x have at most one interaction with matter per orbit. Therefore WIMPs are not really in thermal equilibrium with nuclei. Press and Spergel [5] made the strong assumption that WIMPs form an isothermal gas over their region of extension. They define a constant WIMP temperature T_x which is of the order of the nuclei temperature $T(r)$. At the center of the star T_x is smaller than $T(0)$: so in this region the net effect of a WIMP-nucleus collision is that the WIMP gains energy whereas the nucleus looses some. If after this collision the WIMP goes to the outskirts of the extension region, there T_x is larger than $T(r)$ and the net effect of a WIMP-nucleus collision will be that the WIMP will loose the energy gained in the center of the star. This process is very efficient to evacuate the nuclear energy from the center of a star : it can be called a 'cooling bridge'. The energy $\varepsilon_x(r)$ transferred per unit time and unit mass from nuclei to WIMPs is proportional to [5]

$$n_x(r)\ \frac{1}{l_x(r)}\ (T(r) - T_x)$$

where $n_x(r)$ is the number density of WIMPs at distance r. The WIMP temperature T_x is fixed by the condition of non-evaporation : the total luminosity carried out of the star by WIMPs is zero, i.e. $L_x(R)=0$.

The WIMP 'cooling bridge' effect leads to an isothermalization of the core of radiative stars and consequently to a decrease of the central temperature and of its gradients. It has also an influence on the central convection of convective stars (see talks by P.Salati and G.Raffelt).

The isothermal gas approximation has been criticized by Nauenberg [11] and Gould and Raffelt [10]. They argued that in the core of stars, especially in the sun, even if the temperature gradient is small, nuclei do not constitue an isothermal gas. Moreover the isothermal WIMP gas approximation, i.e. to consider a Maxwellian WIMP distribution, is less and less valid as their velocity becomes larger and larger and approaches the escape velocity. To test the Press and Spergel approximation, both groups did Monte Carlo simulations of the Brownian motion of WIMP in the core of the sun. They concluded that the Press and Spergel formula for the WIMP luminosity $L_{x,PS}(r)$ could be considered as a rather good approximation, provided it is multiplied by a factor 0.5. So in the non local (NL) regime the WIMP luminosity satisfies

$$L_{x,NL}(r) \sim 0.5\ L_{x,PS}(r)$$

4.2 Conductive regime

In this regime the WIMP mean free path is much smaller than the dimension of the region where they are trapped, so it corresponds to $K_n \ll 1$. In practice this regime is effective for $K_n \leq 0.1 - 0.2$. In this regime WIMPs are roughly in local thermal equilibrium with nuclei and therefore it is possible to define a WIMP temperature T_x which equals the nuclei temperature T(r). The WIMP luminosity is proportional to

$$n_x(r) \; l_x(r) \; \left(-\frac{dT(r)}{dr} \right)$$

It is very similar to the photon or radiative luminosity. Gould and Raffelt[10] have solved the Boltzmann equation for WIMP and obtained the conductive luminosity $L_{x,COND}(r)$.

Having noted that for all values of K_n , $L_{x,NL}(r)$ and $L_{x,COND}(r)$ have roughly the same shape with a maximum for $r \sim r_x$ we fit the Gould-Raffelt Monte Carlo results with the following linear combination :

$$L_x(r) = \frac{1}{1 + 1/(5 \; K_n^3)} \; L_{x,NL}(r) \; + \; \frac{1}{1 + 52 \; K_n^3} \; L_{x,COND}(r)$$

This is a reasonable WIMP luminosity valid for any value of K_n ; therefore we plugged it into the Saclay solar evolution code, incorporating, as such, the WIMP energy transport[12].

In the Knudsen regime, when the WIMP-matter cross-section increases (at fixed WIMP mass), the WIMP energy transport is more and more effective and so is the cooling of the star core. As the cross-section keeps on increasing we reach the inter-mediate regime $(1 \geq K_n \geq 0.2)$ in which the conductive energy transport starts to be effective over the non-local one, reducing the cooling effect. Then as the cross-section σ_{xN} still continues to increase we fully enter the conductive regime : the cooling effect decreases until it completely disappears.

For the sun, the cooling effect is effective enough to solve the solar neutrino problem for WIMP mass between 2 GeV and 7 GeV and WIMP-matter cross-section between $10^{-36} \; cm^2$ and $10^{-34} \; cm^2$. Then in this case WIMPs are called cosmions.

5 Cosmion candidates

The problem is that no existing particle theory leads to candidates which enter into the range of mass and cross-section described at the end of last section. Roughly the cosmion-matter cross-section is too strong to be weak and too weak to be strong.

As the de Broglie wave length of cosmions with mass a few GeV and speed around 300 km/s is larger than the size of most nuclei, the scattering of cosmions on nuclei is coherent. Using the cosmion-hydrogen cross-section σ_{xp} as a free parameter, we relate the cross-sections of cosmion on other nuclei to this free parameter in the framework of two different coherent models[13] :

1.Axial coupling model :

$$\sigma_{xN} \simeq \frac{4}{3} \; J(J+1) \left[\frac{A(1 + m_x/m_p)}{A + m_x/m_p} \right]^2 \sigma_{xp}$$

where J is the nuclear spin and A the atomic number. The interaction is spin dependent, the result of this dependence being that in this model cosmion scatters predominantly on hydrogen.

2. Vector interaction model :

$$\sigma_{xN} = A^2 \left[\frac{A(1 + m_x/m_p)}{A + m_x/m_p} \right]^2 \sigma_{xp}$$

Unlike in the previous model, here the interaction is spin independent. Moreover the cross-section grows very fast with the atomic number of the nuclei which leads to the fact that cosmions scatter much more efficiently on helium than on hydrogen. In addition, even if their abundances inside the sun are small, scatterings on nitrogen, oxygen and carbon should not be neglected.

6   Constraints on cosmion parameters

We give results for the cosmion parameter space, in the framework of the Saclay solar evolution code and for the vector interaction model [12]. The cosmion parameter space is represented on figure 1 which relates the cosmion mass to its cross-section on hydrogen. Those cosmions decrease the solar neutrino rate from 5.8 SNU to 2 and 2.5 SNU respectively. For the dark matter halo, we use the following input values : $\rho_{halo} = 0.4$ GeV$/cm^3$ and $v_{halo} = 300$ km/s ; errors are taken into account by a factor 2 on the ratio ρ_{halo}/v_{halo}.

The interest of the vector interaction model is that it leads to cosmion cross-sections which are much larger on heavy nuclei than on hydrogen. The figure shows that the appropriate σ_{xp} range is roughly $10^{-37}cm^2 - 10^{-35}cm^2$. To get the range for the cosmion cross-section on He^4 it is necessary to multiply σ_{xp} by a factor which is approximately 100. For the cosmion cross-section on Si^{28} the factor is approximately 15000, leading to a range : $10^{-33}cm^2 - 10^{-31}cm^2$; and for the cosmion cross-section on Ge^{72} the factor is approximately 120000, leading to a range : $10^{-32}cm^2 - 10^{-30}cm^2$. Present dark matter detectors use a crystal of Germanium or Silicium, and the related cross-sections happen to be in the ranges currently tested (see talks by C.Tao and Y.Giraud-Héraud).

References

[1] R.Davis, D.S.Harmer and K.C.Hoffman : Phys.Rev.Lett. 20 (1968) 1205 ; R.Davis : in "Inside the Sun", IAU Symposium 121, Versailles, France, May 1989, Kluwer Academic Publishers.

[2] J.N.Bahcall, R.K.Ulrich : Rev. Modern Phys. 60 (1988) 297.

[3] S.Turck-Chièze, S.Cahen, M.Cassé and C.Doom : Astrophys. J. 335 (1988) 415.

[4] G.Steigman, C.L.Sarazin, H.Quintana, J.Faulkner : Astrophys. J. 83 (1978) 1050 ; L.M. Krauss : Harvard preprint (1985) HUTP-85/A008a, unpublished ; J.Faulkner and R.L.Gilliland : Astrophys. J. 299 (1985) 994.

[5] D.N.Spergel and W.H.Press : Astrophys. J. 294 (1985) 663.

[6] W.H.Press and D.N.Spergel : Astrophys. J. 296 (1985) 679.

[7] A.Gould : "Evaporation of WIMPs with arbitrary cross-sections", IAS-Preprint AST 89/49, submitted to The Astrophys. J..

[8] A.Bouquet, J.Kaplan and F.Martin : Astron. Astrophys. 222 (1989) 103.

[9] A.Bouquet and P.Salati : Astron. Astrophys. 217 (1989) 270.

[10] A.Gould and G.Raffelt : "Thermal conduction by massive particles", IAS-Preprint AST 89/12, to be published in Astrophys. J. ; "Cosmion energy transfer in stars : the Knudsen limit", IAS-Preprint AST 89/20, to be published in Astrophys. J..

[11] M.Nauenberg : Phys.Rev. D36 (1987) 1080.

[12] Y.Giraud-Héraud, J.Kaplan, F.Martin de Volnay, C.Tao and S.Turck-Chièze : "WIMPs and solar evolution code", to be published in Solar Physics (1990).

[13] M.W.Goodman and E.Witten : Phys.Rev. D31 (1985) 3059 ; A.De Rújula, S.L.Glashow, L.Hall : Nature 320 (1986) 38.
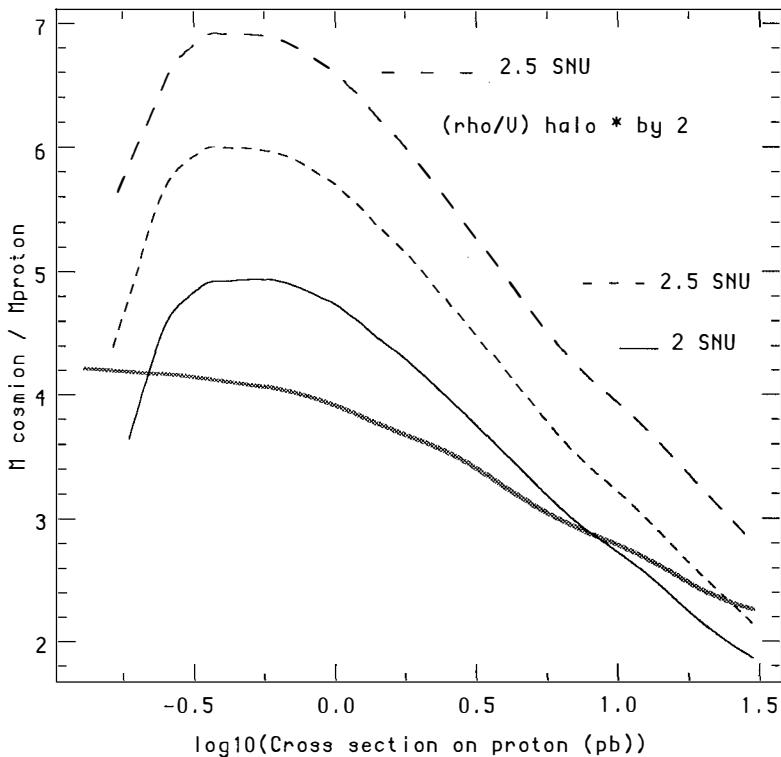
Figure 1: Cosmion parameter space for vector interaction coherent model.The abscissa shows $log_{10}(\sigma_{xp})$ with σ_{xp} expressed in picobarn=$10^{-36}cm^2$, while the ordinate shows m_x/m_p. The lower gray curve is the evaporation curve : the allowed region lies above that curve. The two curves which are just above are respectively the 2 and 2.5 SNU curves obtained for the mean values of ρ_{halo} and v_{halo} (see text). The upper dotted curve is a 2.5 SNU curve whose difference with the one just below lies in the fact that the ratio ρ_{halo}/v_{halo} is multiplied by 2.

Dark matter and stellar evolution

Pierre Salati [a,1] and Alain Bouquet [b]

a) Theory Division, CERN, CH 1211, Genève 23, Switzerland
b) LPTHE Université Paris VII, 75251 Paris Cedex 05, France

Abstract

The existence of weakly interacting massive particles (WIMPs) has been suggested to explain both the dark matter conundrum and the solar neutrino deficiency. Provided that they do not annihilate, these particles can be extremely efficient in transporting energy inside stellar cores. They can supplement radiation and, under the specific circumstances which we analyze, even suppress core convection. We review here the main modifications experienced by stellar evolution when WIMPs are present. We conclude that cold dark matter leads to an anomalous mass-to-luminosity relation for light ($M < 0.5\,M_\odot$) main-sequence stars and also implies thermal pulses in horizontal branch stars.

---

[1] On leave of absence from LAPP, BP110, 74941 Annecy-le-Vieux Cedex, France and Université de Chambéry, 73000 Chambéry, France.

1 - Introduction.

The nature of the halo dark matter is still unresolved (see for instance [1]), but the existence of a new, weakly interacting massive particle (hence the acronym WIMP) is one the most exciting possibilities suggested to solve this problem. These particles may constitute the halo of galaxies and can be accreted by stars [2]. They steadily accumulate inside stellar cores – unless they annihilate by pair – and they can efficiently transport energy, competing with radiative transport or convection. A lot of effort is being devoted to detect directly these particles which, through their collisions with nuclei, might deposit energy inside terrestrial bolometers or let visible ionization tracks inside semi-conductors. Another promising approach is the study of the possible effects of these dark matter particles on stellar evolution, and the search for any observational anomaly which would indirectly sign their presence inside stellar cores. In this contribution, we try to indicate the regions of the Hertzsprung-Russel diagram where WIMPs are expected to be most effective.

As recalled by F. de Volnay, WIMPs supplement stars with an additional mechanism of energy transport and tend to decrease their temperature gradient. Stars such as red-giants or white dwarfs which already possess isothermal interiors are not expected to be affected by WIMPs. More sensitive to these particles are presumably stars with radiative cores. Because WIMPs may decrease the central temperature of the Sun, they have been advocated as a possible solution to the solar neutrino deficiency [3]. The Sun and other $\sim 1~M_\odot$ stars in which the central energy transport is enhanced by WIMPs may leave the main-sequence earlier than what is conventionally acknowledged. The age of the models reaching the turn-off surface temperature may be reduced by the WIMPs by a factor 15-20% [4]. The most dramatic effect of dark matter particles is the suppression of core convection inside main-sequence and horizontal branch (HB) stars. Convection constantly replenishes the central burning region with fresh fuel, and its suppression is therefore expected to decrease stellar lifetimes, hence a modified stellar evolution. We give in section 2 the criterion under which central convection is suppressed by dark matter, and analyze it for the main-sequence. Section 3 is devoted to an analytic discussion of the thermal pulses inside the helium cores of HB stars, induced by a low central opacity – a typical situation in the presence of WIMPs.

While WIMPs may eventually fade away, note that our discussion of low opacities and stellar evolution is fairly generic. Trying to detect cold dark matter through its indirect signatures on stars may eventually lead to a better understanding of their inner thermodynamics. Stars appear actually to be much more active and, for that matter thermal pulses much more frequent, than what was previously acknowledged.

2 - The main-sequence.

Stars are powered by nuclear reactions. In Figure 1, the energy production rate ϵ_{nuc} at the center of main-sequence stars is plotted as a function of the stellar mass M [5]. For $M < 1~M_\odot$, the pp chain dominates while for heavier stars, the CNO cycle takes over, hence a steeper slope for ϵ_{nuc}.

In a radiative core, the temperature gradient adjusts itself so that the nuclear energy is entirely carried away by radiation. However, the temperature gradient cannot exceed its adiabatic value, otherwise convection develops and transports the excess energy. Radiative transport is therefore most
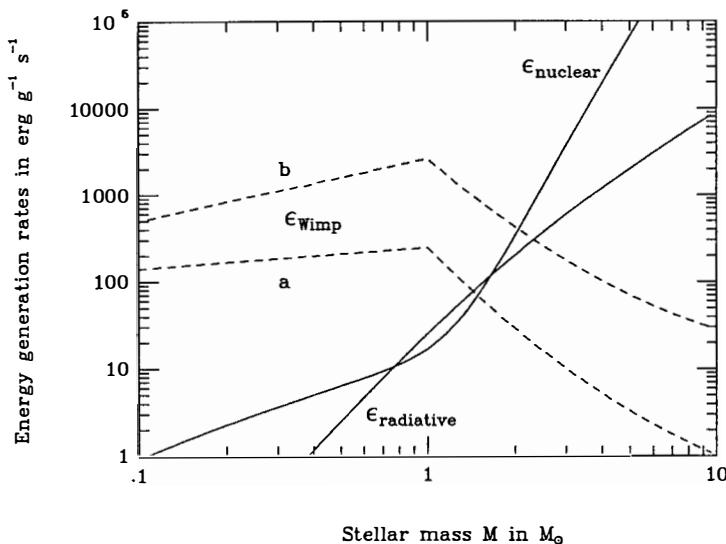
Figure 1

Nuclear energy production rate ϵ_{nuc} and the maximal radiative transfer rate ϵ_{rad} at the center of main-sequence stars are plotted as functions of the stellar mass M (solid lines), and compared to the rate of WIMP energy transfer ϵ_{Wimp} (dashed lines). Curve **a** corresponds to a WIMP of mass $m_W = 4\,m_p$ and effective cross-section of 5×10^{-37} cm^2. Curve **b** corresponds to a mass $m_W = 12\,m_p$ and a cross-section of 2×10^{-35} cm^2. Any realistic WIMP explanation of the solar neutrino deficiency would lie in between and would lead to the suppression of core convection inside low-mass stars.

efficient whenever the temperature gradient is adiabatic, so that the maximal amount of energy which photons can withdraw from the stellar center is given by :

$$\epsilon_{rad} = \frac{128\pi}{15} \frac{\sigma_{Stefan}}{\kappa(0)\rho(0)} \frac{G\mu m_p}{k} T^3(0) \ . \tag{1}$$

σ_{Stefan} is the Stefan-Boltzmann constant, $\kappa(0)$, $\rho(0)$ and $T(0)$ are respectively the central opacity, density and temperature whilst μ is the mean molecular weight in units of the proton mass m_p. The evolution of ϵ_{rad} with M is also displayed in Figure 1. As discussed by Schwarzschild, the onset of convection in stellar cores arises whenever $\epsilon_{nuclear}$ exceeds ϵ_{rad} :

• For stars in the range $0.7\,M_\odot < M < 1.7\,M_\odot$, cores are radiative – a situation recovered in the case of the sun by numerical simulations.

• Heavy stars, *i.e.*, $M > 1.7\,M_\odot$, have convective cores since the CNO cycle produces far more energy than what photons are able to transport. On the other hand, it is the increase of the central opacity of low mass stars, *i.e.*, $M < 0.7\,M_\odot$, which makes them convective all the way down to the center.

Dark matter particles are accreted during the main-sequence stage so that, depending on their effective cross-section σ and mass m_W, they may dominate the energy transport inside stellar cores. If their effective energy transfer rate ϵ_{Wimp} is large enough, core convection is no longer necessary since dark matter particles and photons are able to carry the bulk of the nuclear energy. The criterion for core convection suppression by WIMPs is therefore :

$$\epsilon_{rad} + \epsilon_{Wimp} \geq \epsilon_{nuc} \geq \epsilon_{rad} \ . \tag{2}$$

The two dotted curves of Figure 1 display the evolution of ϵ_{Wimp} with M for :
(a) $m_W = 4\,m_p$ and $\sigma = 5 \times 10^{-37}$ cm^2 and
(b) $m_W = 12\,m_p$ and $\sigma = 2 \times 10^{-35}$ cm^2.
Both curves exhibit a kink around $1\,M_\odot$ and drop sharply for heavy stars which lifetime along the main-sequence is too short for significant WIMP accretion. On the other hand, ϵ_{Wimp} exceeds ϵ_{nuc} by at least two orders of magnitude in the case of low-mass stars – $M < 0.7\,M_\odot$ – which core convection is therefore suppressed. Note that any realistic WIMP explanation of the solar neutrino deficiency would lie between curves (a) and (b), and would inevitably lead to the suppression of core convection. Future detailed numerical investigations should focus on stars in the mass range $0.2 - 0.5\,M_\odot$ and explore the modifications in their structure implied by the presence of dark matter. We crudely guess that the core readjustment, after central convection is suppressed, implies an increase by a factor ~ 3-4 of the total stellar luminosity and generates an inflation of the stellar radius by $\sim 50\%$. Binaries for which the mass, luminosity and radius of each component are well determined should be scrutinized since any departure from the conventional mass-radius relation for one of the stars would sign the suppression of core convection and, therefore, the presence of WIMPs.

3 - Horizontal branch stars.

Core convection also plays a crucial role inside HB stars since it replenishes the small nuclear core with the fuel from a significantly more extended region. If WIMPs are present inside HB cores, criterion (2) is fulfilled so that dark matter particles may inhibit core convection on the horizontal branch [6].

The helium core of a typical HB star may be divided into two zones [7]. The first is the central burning region with mass $M_1 = 0.05\,M_\odot$, temperature $T_1 \approx 10^8\,K$ and a positive heat capacity c_1. The nuclear energy generation rate ϵ_1 varies as $\sim T_1^{40}$. The second zone is the remainder of the core with mass $M_2 \approx 0.5\,M_\odot$ and constitutes a gravo-thermal buffer with negative heat capacity $-c_2$. The temperature evolution in the two regions is modelled by a set of non-linear differential equations :

$$
\begin{aligned}
\frac{dT_1}{dt} &= F_1(T_1, T_2) = \left(\frac{1}{c_{p1}M_1}\right)\left\{\epsilon_1 M_1 - K_1\left(T_1^4 - T_2^4\right)\right\}, \\
\frac{dT_2}{dt} &= F_2(T_1, T_2) = \left(\frac{-1}{c_2 M_2}\right)\left\{K_1\left(T_1^4 - T_2^4\right) - K_2\left(T_2^4 - T_{ext}^4\right)\right\}.
\end{aligned}
\tag{3}
$$

where T_{ext} is the temperature outside of the core, K_1 accounts for the energy transfer between regions 1 and 2 by means of radiative and WIMP transfer, and K_2 is the thermal conductance between the core and the envelope. This set of differential equations is supplemented by the condition that when $T_1 - T_2$ exceeds a critical value $\nabla_{ad}T \simeq 2.59 \times 10^7$ K, convection transports energy so efficiently that the temperature difference remains fixed at this value and the nuclear burning energy is directly dumped into the gravo-thermal buffer. This simple model is sufficient to account for the numerical results which G. Raffelt will subsequently discuss. The thermal conductance $K_1 \propto 1/\kappa_1$ so that the core evolution is completely specified by the opacity at the center of the HB star where exotic phenomena (such as the presence of WIMPs) may increase the energy transport.

- $\kappa_1 > \kappa_{crit}$: if K_1 is chosen too small to break convection, equation (3) has a fixed point $E_{conv} \equiv (T_{1\,conv}, T_{2\,conv})$ corresponding to the usual convective equilibrium solution of stellar structure. The existence and location of this solution does not depend on the assumed value of the central opacity, κ_1, while its stability does depend on the choice of κ_1.

- $\kappa_1 < \kappa_{crit}$: if K_1 is so large that convection is broken, equation (3) has a limit cycle : $[T_1(t), T_2(t)]_{limit}$, shown in the T_1-T_2-plane of Figure 2. Note that the dashed diagonal line represents the condition for convection, $T_1 - T_2 = \nabla_{ad}T$, so that the area below is unphysical since over-adiabaticity is negligible. Any evolution in the presence of convection occurs therefore along this boundary.

For a standard central opacity, $\kappa_1 = 0.2\,\mathrm{cm^2\,g^{-1}}$, the state E_2 of conductive equilibrium falls into the physically forbidden regime of over-adiabaticity. For the critical value of the central opacity,

$$
\kappa_{crit} = 3.1 \times 10^{-2}\,\mathrm{cm^2\,g^{-1}}, \tag{4}
$$

it lies on the convection line and coincides with E_{conv} while, for $\kappa_1 < \kappa_{crit}$, it lies well above the convective border, at the crossing between curves (1) ($\dot{T}_1 = 0$) and (2) ($\dot{T}_2 = 0$). For $\kappa_1 = 0.01\,\mathrm{cm^2\,g^{-1}} < \kappa_{crit}$, both states E_{conv} and E_2 are unstable and do not correspond to stationary physical configurations of the star. The convective equilibrium point lies in the region of the T_1-T_2-plane where $\dot{T}_1 < 0$, i.e., nuclear burning is quenched by the suppression of convection in agreement with Schwarzschild's criterion (2). The conductive equilibrium point lies on the unstable portion of curve (1) where T_2 drops sharply with increasing T_1 : if the star moves slightly to the left of this point, nuclear burning gets quenched, if it moves slightly to the right, nuclear burning increases, and the nuclear core runs thermally away until the trajectory is "intercepted" by the convection line.
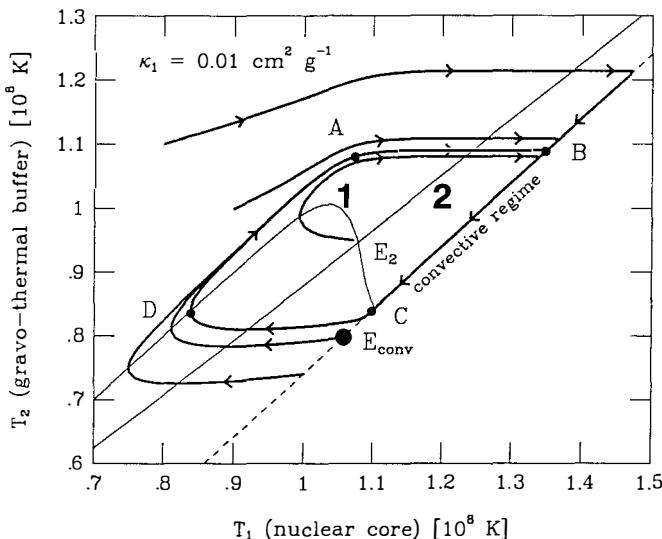
Figure 2

Several explicit trajectories for the pulsating star in the T_1-T_2-plane, with $\kappa_1 = 0.01\,\mathrm{cm}^2\,\mathrm{g}^{-1} < \kappa_{\mathrm{crit}}$. All trajectories approach the limit cycle.

The nature of the limit cycle is best understood if we follow one period of the pulsation explicitly. The variation of various stellar parameters along the cycle are presented in Figure 3. The nuclear luminosity, $L_{3\alpha} = \epsilon_1 M_1$, denotes the energy generated by helium burning at the center of the star. Part of this luminosity goes into expansion work of the gravo-thermal buffer and is stored as gravitational energy. The actual luminosity which escapes the helium core is given by $L_{\mathrm{core}} \equiv \mathrm{K}_2 (T_2^4 - T_{\mathrm{ext}}^4)$. The upper panel presents the evolutions of L_{core} (solid curve) and $L_{3\alpha}$ (dashed curve) during a pulse. The variation of the nuclear core temperature, T_1, is displayed in the intermediate panel, and the last diagram indicates which part of the cycle corresponds to the presence of convection.

a) The quiescent contraction (DA)

We begin at point D which is the intersection of the limit cycle with curve (1). The nuclear activity of the central region is negligible, and the temperature of the nuclear core closely follows the gravo-
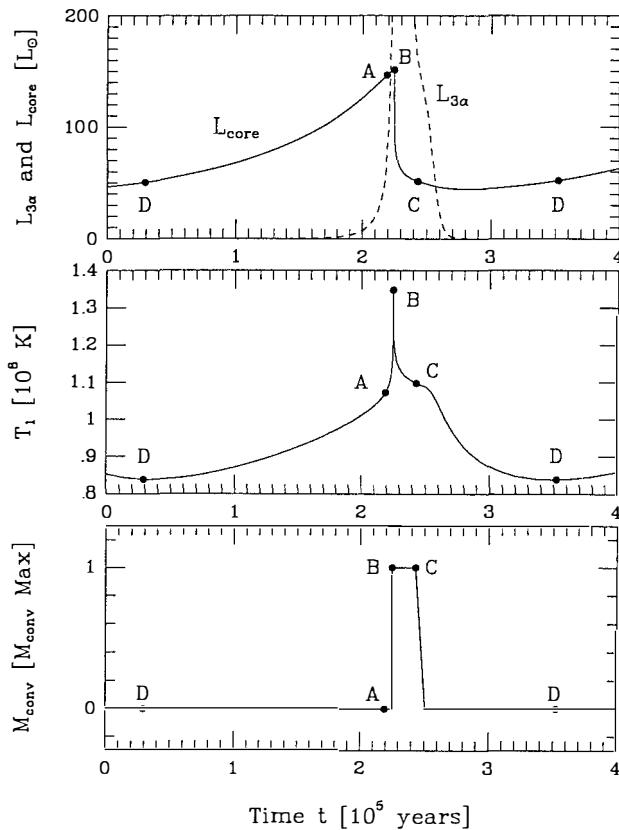
Figure 3

Evolution of the stellar parameters for one oscillation period according to our analytical model with $\kappa_1 = 0.01\,\mathrm{cm^2\,g^{-1}} < \kappa_{crit}$. Upper panel: Core luminosity, L_{core}, (solid curve), and nuclear energy production, $L_{3\alpha}$, (dashed curve). Intermediate panel: Central temperature, T_1. Lower panel: The curve takes the value 0 in the absence of convection, and 1 in its presence. The points A, B, C, and D correspond to the points shown on the limit cycle of Figure 2.

thermal buffer, $T_1 \sim T_2$, and $\dot{T}_1 \sim \dot{T}_2$. The evolution is dominated by the gravo-thermal buffer which slowly contracts on the Kelvin-Helmholtz timescale which is determined by its heat capacity and the thermal conductance to the envelope, K_2. The nuclear core can be thought of as being just a small part of the gravo-thermal buffer. The core luminosity, L_{core}, is provided by gravitational energy liberated because of the contraction.

b) The violent nuclear eruption (AB)

In point A, the core becomes unstable because the central temperature has reached the critical point where nuclear burning comes to life. Conduction is no longer able to cool the nuclear core efficiently enough and it runs thermally away : its temperature, T_1, almost instantaneously jumps to such a large value that convection sets in (point B). The temperature of the gravo-thermal buffer remains almost unchanged, the evolution is practically horizontal in the T_1-T_2-plane: the evolution of the two zones is virtually independent of each other. The exceedingly short time scale of this phase is given by the steep temperature variation of the triple alpha reaction, and by the small heat capacity of the nuclear core. The central temperature, T_1, jumps brutally from $\sim 1.05 \times 10^8$ K to $\sim 1.4 \times 10^8$ K and causes helium combustion to erupt violently by several orders of magnitude.

c) The convective period (BC)

After reaching point B, convection commences, pouring the nuclear energy directly into the gravo-thermal buffer. Nuclear reactions are extremely active on the path from B to C, but the peak central temperature was reached at point B. The moderating effect of the buffer becomes now fully operative as the continuous energy production leads to expansion and cooling. The action of the two regions is anti-thetical, the nuclear core produces energy and would lead to further heating, while the expansion of the buffer absorbs the energy and cools. The time-scale of this phase is determined by a combination of the properties of the nuclear core and the gravo-thermal buffer, *i.e.*, by the energy generation rate of the nuclear core and by the (negative) heat capacity of the buffer. This time scale is much larger than that of the preceding eruption which depended only on the (positive) heat capacity of the nuclear core which is much smaller because of its ten times smaller mass. Since the temperature of the buffer, T_2, decreases during this phase, the overall luminosity, L_{core}, also decreases as shown in the upper panel of Figure 3 (solid line).

d) The sudden nuclear freeze (CD)

In point C, convection ceases to operate, and the central energy transfer proceeds again by conduction. As the temperature of the inner core falls, the nuclear activity decreases, accelerating the process of cooling because of the small heat capacity of this region. In other words, the nuclear core now "runs away" in the opposite direction compared to AB, and nuclear burning is quenched as can be seen on the upper panel of Figure 3. After this sudden nuclear freeze, the star quickly reaches point D and begins its quiescent contraction.

4   - Conclusions.

WIMPs, and for that matter any exotic mechanism of energy transport at the center of stars, may induce significant observational effects on stellar evolution, affecting specific regions of the Hertzsprung-Russel diagram :

• On the main-sequence, low mass stars have an anomalous mass-to-radius relation which translates into an increase of their luminosity by $\sim$ a magnitude.

• Radiative stars in the range 0.7 $M_\odot - 1.7 M_\odot$, have an isothermal core which implies, according to the Schoenberg-Chandrashekar's theorem, a shorter lifetime on the main-sequence.

• This opens the question of the sub-giant branch where fast-evolved stars from the main-sequence gather before expanding as red-giants. There is either an excess of these stars in the sub-giant population if helium core contraction towards a degenerate state requires complete hydrogen exhaustion, or the inferred age of globular clusters is shorter than what is conventionally acknowledged on the basis of standard stellar evolution along the main-sequence.

• Core convection is suppressed in stars where the nuclear energy production rate is fairly temperature sensitive. The star does not settle down in a steady state of stable conductive equilibrium but undergoes thermal pulses. This is the case for HB stars and is also presumably true for $\sim 2 M_\odot$ main-sequence objects inside which WIMP conduction is still fairly effective and where the CNO cycle depends sensitively on the temperature.

References

[1] J. Kormendy and G. R. Knapp; in *IAU Symposium 117, Dark Matter in the Universe*, eds. J. Kormendy and G. R. Knapp (1986) (Dordrecht : Reidel)

[2] W. H. Press and D. N. Spergel; *Ap. J.* **296** (1985) 679.

[3] G. Steigman, C. L. Sarazin, H. Quintana and J. Faulkner; *Ap. J.* **83** (1978) 1050: D. N. Spergel and W. H. Press; *Ap. J.* **294** (1985) 663; L. M. Krauss; preprint HUTP-85/A008a (1985); J. Faulkner and R. L. Gilliland; *Ap. J.* **299** (1985) 994.

[4] J. Faulkner and F. J. Swenson; *Ap. J.* **329** (1988) L47.

[5] A. Bouquet and P. Salati; *Ap. J.* **346** (1989) 284.

[6] A. Renzini; *Astr. Ap.* **171** (1987) 121.

[7] D. Dearborn, G. Raffelt, P. Salati, J. Silk and A. Bouquet: *Nature* **343** (1990) 347; preprint CfPA-TH-89-009 (1989); P. Salati, G. Raffelt and D. Dearborn; preprint CfPA-TH-89-012 (1989).

DARK MATTER AND THERMAL PULSES IN HORIZONTAL BRANCH STARS

Georg G. Raffelt

Max-Planck-Institut für Physik

Postfach 401212, D-8000 München 40, West Germany

ABSTRACT

We discuss the effect of the energy transfer by dark matter (DM) particles on horizontal branch (HB) stars, emphasizing numerical results and observational consequences. For a DM density similar to that of the solar neighborhood, HB stars would contain enough DM particles to break core convection, leading to thermal pulses on a time scale of $\sim 5 \times 10^5$ yrs. The overall duration of the HB phase decreases and the luminosity dispersion of the HB increases, but neither effect is pronounced enough to conflict with observations. The magnitude difference between the HB and the main sequence turnoff increases and leads to an overestimate of globular cluster ages. The observed period changes of RR Lyrae stars are consistent with, and even implied by, our scenario.

INTRODUCTION

Following two talks concerning the effects of DM particles on stars, I can keep my introduction extremely brief. The main idea is that DM particles will be trapped in stars[1], build up there over the stellar lifetime (assuming they do not annihilate or evaporate), and will contribute to the transfer of energy in the central regions. For suitable masses and scattering cross sections, the central temperature of the Sun can be reduced enough to solve the solar neutrino problem[2,3], in which case our particles are termed "cosmions". (The relevant parameter range was discussed in the preceding talk by F. Martin de Volnay.) Moreover, they could break core convection in HB stars[4], an effect that would lead to thermal pulses rather than to an overall reduction of the helium burning lifetime as was shown in a paper by Dearborn, Raffelt, Salati, Silk, and Bouquet[5] on which this lecture is based. An analytic explanation for this behavior was put forth by Salati, Raffelt, and Dearborn[6], and was discussed in the preceding talk by P. Salati. I want to discuss these effects on HB stars, emphasizing numerical results and observational consequences.

WHICH PARTICLE PROPERTIES?

If the DM particles are supposed to transfer energy in the cores of HB stars which consist mainly of ^{4}He, ^{12}C, and ^{16}O, they must interact with these spin-zero nuclei. Hence it is convenient to use the scattering cross section on helium, σ_{He}, to parametrize the interaction. We will use a coherent interaction law where the cross-section of a cosmion with "atomic weight" A_x on a nucleus of atomic weight A_j is given by

$$\sigma_j = \sigma_{\mathrm{He}} \left(\frac{A_x + 4}{A_x + A_j} \right)^2 \left(\frac{A_j}{4} \right)^4 . \tag{1}$$

Moreover, we must require that our particles neither annihilate nor evaporate during the HB lifetime or during an earlier phase of evolution, a requirement which also implies that the total number of DM particles in the star is given by the capture on the main sequence (MS) progenitor,[1,7]

$$N_x = 2.5 \times 10^{47} \frac{1}{A_x} f_{\mathrm{MS}} \left[1 - \exp\left\{ -\frac{\sigma_{\mathrm{cap}}}{\sigma_{\mathrm{crit}}} \right\} \right] , \tag{2}$$

where m_x is the particle mass. The factor f_{MS} characterizes the MS progenitor and its DM environment by

$$f_{\mathrm{MS}} \equiv \left(\frac{\rho_{\mathrm{DM}}}{0.01 \, M_\odot/\mathrm{pc}^3} \right) \left(\frac{300 \mathrm{km/s}}{\overline{v}} \right) \left(\frac{M}{0.8 \, M_\odot} \right)^{1.8} \left(\frac{t_{\mathrm{MS}}}{10^{10} \, \mathrm{yr}} \right) , \tag{3}$$

where ρ_{DM} is the dark matter density, $\overline{v}$ its velocity dispersion, M the progenitor mass, and t_{MS} the main sequence lifetime, and we have assumed that the radius of the pro-

genitor varies as $M^{0.8}$. Our benchmark values for ρ_{DM} and $\bar{v}$ characterize the solar neighborhood, and the mass $M = 0.8\,M_\odot$ corresponds to an evolved globular cluster star today. The capture cross section, σ_{cap}, is a weighted average of the scattering cross sections on hydrogen and helium which are the major constituents of a MS star, $\sigma_{cap} \sim 0.9\,\sigma_H + 0.1\,\sigma_{He}$, and $\sigma_{crit} = 3{\times}10^{-36}\,cm^2$ is the cross section where capture saturates, i.e., every particle impinging on the star is captured.

As a further assumption we require that σ_{He} is so large that the energy transfer in the core of an HB star proceeds by conduction rather than by the non-local mechanism discussed in the previous talks. In this case the efficiency of energy transfer scales with N_x/σ_{He} which is the relevant figure of merit for our discussion. In the conduction limit, the energy transfer can be treated exactly, in particular because the relevant thermal condution coefficient has recently been determined[8]. For details we refer to Refs. 8 and 5.
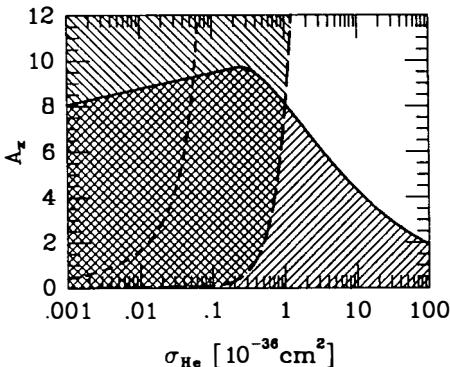
Figure 1: The particle parameters must lie above the "evaporation line" (solid curve) in order to be retained by the HB star over its lifetime of $\sim 10^8$ yr, and to the right of the long-dashed curve so that energy transfer proceeds by conduction. The short-dashed line gives the locus of parameters for Dirac neutrinos. A_x is the particle "atomic weight", i.e., its mass in GeV.

In Fig. 1 we show the relevant range for mass, m_x, and cross section, σ_{He}, which we consider. Our DM particles must lie in the white area: above the "evaporation line" below which they would not be ratained by the HB star during its $\sim 10^8$ yr lifetime, and to the right of the "conduction line". For comparison we also show, as a short-dashed line, the locus of parameters for a Dirac neutrino. This means that we require cross sections substantially in excess of standard weak magnitude, and it is this requirement

which constitutes the major (possibly fatal) problem of the entire cosmion scenario. If the cross section on hydrogen is negligible, our particles solve the solar neutrino problem for σ_{He} around a few tens of 10^{-36} cm^2, *i.e.*, our parameter range encompasses cosmions. In other words, if cosmions solve the solar neutrino problem, one naturally may expect convection breaking in HB stars. Hence, because the cosmion interpretation of the solar neutrino problem is still the only, albeit very tentative, signature for the existence of *particle* DM, it is worth investigating other observational consequences of this hypothesis.
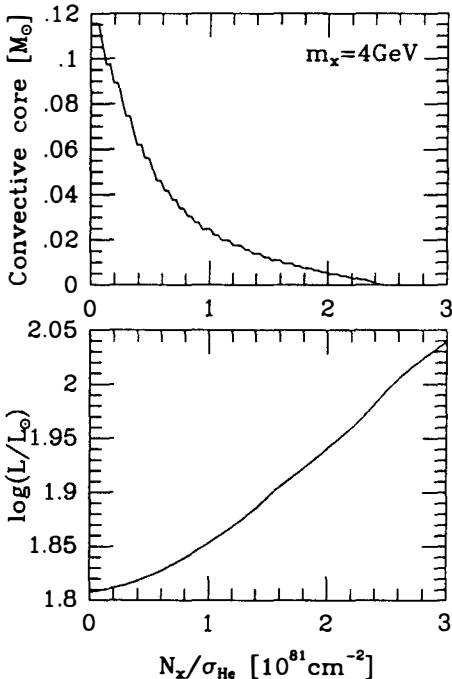
Figure 2: Convective core and luminosity of zero age HB models as a function of N_x/σ_{He} for a particle mass of 4 GeV.

CONVECTION BREAKING AND THERMAL PULSES

A standard HB star consists of a helium core of $\sim 0.5\,M_\odot$ and an extended envelope. At the interface between core and envelope, hydrogen burns in a thin shell, providing the major part of the total stellar energy output. In the center of the core, helium

burns in a small region which is surrounded by a convective core which encompasses initially about $0.12\,M_\odot$. In order to study the effect of DM particles we took a zero age standard HB model and slowly increased the effect of the exotic transfer of energy, i.e., slowly increased N_x/σ_{He} while preventing the star from evolving by keeping the composition profile fixed. In Fig. 2 we show the resulting reduction of the convective core and increase in surface brightness. For $N_x/\sigma_{\mathrm{He}} = 2.4{\times}10^{81}\,\mathrm{cm}^{-2}$ the convective core vanishes entirely, while the brightness has increased by almost 0.5 mag. A similar run with a larger particle mass, $m_x = 8\,\mathrm{GeV}$, yields a much more sudden transition to full convection breaking which occurs for $N_x/\sigma_{\mathrm{He}} = 1.1{\times}10^{81}\,\mathrm{cm}^{-2}$.

For $\sigma_{\mathrm{He}} \lesssim 30{\times}10^{-36}\,\mathrm{cm}^2$ capture on the MS is not saturated so that $N_x \propto \sigma_{\mathrm{He}}$. Expanding Eq. (2) yields $N_x/\sigma_{\mathrm{He}} \sim 1{\times}10^{82}\,\mathrm{cm}^{-2}\,f_{\mathrm{MS}}/A_x$, i.e., convection breaking does not depend on the chosen value for σ_{He}, but depends only on the factor f_{MS}. Apparently convection breaking occurs for $f_{\mathrm{MS}} \sim 1$, i.e., for a dark matter density similar to that of the solar neighborhood. In other words, if cosmions solve the solar neutrino problem, it is natural to expect convection breaking in HB stars.

Next we allowed a stellar model ($M = 0.8\,M_\odot$, $Z = 0.004$) to evolve with particles of mass $m_x = 4\,\mathrm{GeV}$ and an abundance of $N_x = 3.3{\times}10^{81}\,\mathrm{cm}^{-2}$ to evolve from the zero age HB to the asymptotic giant branch (AGB). The result of this run is shown in Fig. 3 where it is compared with the standard evolution. The most dramatic feature are thermal pulses on the Kelvin-Helmholtz time scale of the core of $\sim 5{\times}10^5$ yr which were analytically interpreted in the preceding talk by P. Salati. Because of the short bursts of convection which are associated with these pulses, the total supply of helium for burning in the center is not much different from the standard case and the duration of the HB is not changed much.

As helium burns to carbon and oxygen, the total efficiency of our particles to transport energy decreases because of the assumed coherent interaction law. Therefore, after a sufficient amount of carbon has built up in the central region, convection can no longer be broken and the star establishes a permanent convective core, although residual pulses pertain (see Fig. 3).

COMPARISON WITH OBSERVATIONS

One might naively expect that a behavior so dramatic as these thermal pulses should be very apparent observationally. However, while the Kelvin-Helmholtz time scale of the core is short relative to stellar evolutionary scales, it is long in terms of a human life span and thus one has to appeal to statistical methods in order to discuss possible observables. From the upper panel of Fig. 3 it is apparent that the brightness dispersion of HB stars will be increased so that the HB stars in any given globular cluster should
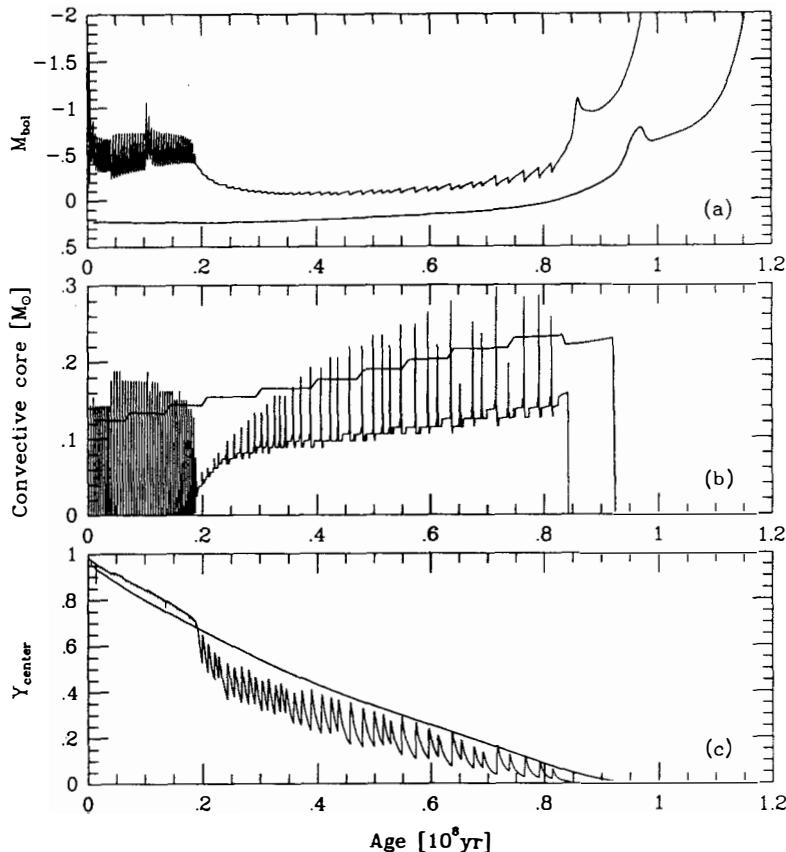
Figure 3: Evolution of a HB star ($M = 0.8 M_\odot$, $Z = 0.004$) with DM particles ($m_x = 4\,\mathrm{GeV}$, $N_x/\sigma_{\mathrm{He}} = 3.3 \times 10^{81}\,\mathrm{cm}^{-2}$) which follow a coherent interaction law. **(a)** Surface brightness *vs.* time where the lower, smooth curve is a standard model without cosmions. **(b)** Size of the convective core where the smooth function is the standard case. The steps in this curve reflect the mass shells on the computer and are of no physical significance. **(c)** Central helium abundance.

occupy a larger range of luminosities than expected on the basis of standard evolutionary calculations. However, if the pulsing phase lasts for only a relatively short fraction of the entire HB evolution as in the run of Fig. 3, this effect is not pronounced enough to be observed.

A more promising observable is the average luminosity of the HB which is shifted to

larger values in spite of the rather short pulsing episode. For the run shown in Fig. 3 the average increase of the HB brightness is about 0.3 mag. If this effect occurs in nature, globular clusters must be about 15% more distant (for the example of Fig. 3) than had been thought previously, an increase that can hardly be excluded on the basis of other evidence. Conversely, the distance of globular clusters can be gauged by comparing the apparent brightness of the HB with theoretical values. These distance determinations then allow one to derive the absolute brightness of the observed MS turnoff which is a measure for the age of the cluster. In other words, the brightness difference between the HB and the MS turnoff is a direct measure, intrinsic to the observed color magnitude diagram, for the age of a given cluster. According to a result of Iben and Renzini[9], an anomalous brightening of the HB by 0.3 mag would correspond to a decrease of the inferred cluster age by $\sim$ 25%. Therefore, the effect of cosmions could help explain the notorious age problem of globular clusters which are difficult to reconcile with the standard cosmological parameters of the Friedmann universe, and which far exceed the age of the galactic disk as inferred from the white dwarf luminosity function.

In a certain region of luminosities and surface temperatures, HB stars exhibit dynamical oscillations and are identified with RR Lyrae stars. These oscillations must not be confused with our thermal pulses on the Kelvin-Helmholtz time scale: the "clock" for these dynamical oscillations is set by the sound travel time from the surface to the center of the star, the period being between several hours and a day. During the pulsing phase, our stars would cross the instability strip and would exhibit a variable dynamical period with a time scale of change on the Kelvin-Helmholtz time scale. It is interesting that the periods of many RR Lyrae stars do, indeed, fluctuate on such time scales! (For a more conventional but somewhat ad hoc explanation of this effect and references to the observations see Ref. 10.)

CONCLUSIONS

We started this investigation thinking that the effect of particle DM on HB stars would exclude a solution of the solar neutrino problem by this mechanism. Our actual findings indicate, on the contrary, that cosmions might kill four birds with one stone: they could constitute the elusive DM of the galaxy, they could solve the solar neutrino problem, they could explain the age problem of globular clusters, and they would explain the observed RR Lyrae period variations! It is unfortunate that one needs to postulate particle properties (large scattering cross sections, small annihilation cross sections) which are hard to produce even by heroic model builders in particle physics, and which are hard to reconcile with the recent LEP constraints on such models (see the contribution of L. Krauss to these proceedings).

290

REFERENCES

1) W. H. Press and D. N. Spergel, *Astrophys. J.* **296** (1985) 679.
2) G. Steigman, C. L. Sarazin, H. Quintana, and J. Faulkner, *Astron. J.* **83** (1978) 1050.
3) D. N. Spergel and W. H. Press, *Astrophys. J.* **294** (1985) 663.
4) A. Renzini, *Astron. Astrophys.* **171** (1987) 121.
5) D. Dearborn, G. Raffelt, P. Salati, J. Silk, and A. Bouquet, *Astrophys. J.*, in press (1990).
6) P. Salati, G. Raffelt, and D. Dearborn, *Astrophys. J.*, in press (1990).
7) A. Bouquet and P. Salati, *Astron. Astrophys.* **217** (1989) 270.
8) A. Gould and G. Raffelt, *Astrophys. J.*, in press (1990).
9) I. Iben Jr. and A. Renzini, *Phys. Rep.* **105** (1984) 329.
10) A. V. Sweigart and A. Renzini, *Astron. Astrophys.* **71** (1979) 66.

IS $M_\nu \simeq 30\,\text{eV}$ NEUTRINO
DARK MATTER (NDM) RULED OUT?

DAVID B. CLINE

Departments of Physics & Astronomy, University of California at Los Angeles
405 Hilgard Avenue, Los Angeles, California 90024

ABSTRACT

We review the possibility that μ or τ neutrinos provide the dark matter to give $\Omega_\circ \simeq 1$. While there is strong theoretical prejudice against this hypothesis, we show that there is no observational evidence that contradicts the possibility of neutrino matter closing the universe. The possible techniques to detect NDM with terrestrial and supernova detectors is reviewed.

I. INTRODUCTION

The issue of what makes the dark matter in the universe is perhaps the most important in astroparticle physics today. While most proponents favor cold dark matter such as super symmetric particle it is still possible that hot dark matter such as massive neutrinos (ν_μ or ν_τ) provides the matter that gives $\Omega = 1$. This matter, being more spread out would be probed less by rotation curves of galactic clusters. However, it is possible to get flat velocity curves in models with neutrinos as the main contribution to Ω as shown in Figure 1.[1]

There is considerable evidence for dark matter in the universe and particle physics inspired models, along with inflation, suggest that $\Omega \simeq 1$, where as there are strong constraints on the amount of baryonic matter in the universe from nucleosynthesis calculations that give

$$\Omega_b \leq 0.15$$

with $\Omega_{\text{luminous matter}} < 0.01$ [1],[2,3]

One or more stable neutrinos, with a mass in the region of 30 eV, could supply the missing or dark matter. The mass relationship is $\Omega_\nu = 1$ for $M_{\nu_x} = 92\, h^2 eV$, where H is the Hubble Constant (h = 1 for H = 100 km/s^{-1}MPC^{-1}; for h = 1/2, $M_{\nu_x} = 23$ eV to give closure of the Universe and for h = 0.6, $M_{\nu_x} = 33$ eV). We thus consider the neutrino mass range of 10 – 40 eV to be of cosmological significance. There are no known laboratory techniques to uniquely detect such a mass directly. It is possible that some form of neutrino oscillation experiment could be used to infer a mass in this range, however, this will depend on the uncertain level of neutrino mixing. The only technique that is known to provide a unique mass measurement is to use the difference in flight times for neutrinos from a distant supernova that goes as

$$\Delta t = 51.4\, R_{\text{MPC}} \left[\left(\frac{M_{\nu_x}}{E_{\nu_x}} \right)^2 - \left(\frac{M_{\nu_e}}{E_{\nu_e}} \right)^2 \right]\ \text{sec}$$

where M_{ν_x} and M_{ν_e} are measured in electro volts, E_{ν_x}, E_{ν_e} are measured in MeV and R (the distance to the supernova) is measured in Mega Parsecs. The μ and τ neutrinos are expected to have higher average energies since they escape from deep inside the supernova core.

Figure 2 shows the expected luminosity function and mean energy of the ν_e, $\bar{\nu}_e$, $\overset{(-)}{\nu_\mu}$, $\overset{(-)}{\nu_\tau}$ neutrinos from the collapse. The important times are:

    (i)   Prompt ν_e $\sim (3-6)\,\text{ms}$
                     Burst

(ii)	Rise time and time internal to acretion pulse	$\sim (100 - 200)\,\text{ms}$
(iii)	Width of the acretion pulse	$\sim 400\text{ms}$
(iv)	Explosion starts	$\sim (300 - 600)\,\text{ms}$
(v)	Start of neutrino cooling	$\sim (300 - 600)\,\text{ms}$
(vi)	Full width of cool-down	$\sim (10 - 20)\,\text{sec}$

These times set the scale of the techniques that may be used to detect a finite neutrino mass by the time of flight method. In Figure 3 we show the difference between a "time of flight" mass measurement at an accelerator and using a supernova.

We assume an instantaneous source of ν_x neutrinos with a distribution of the form[6,7]

$$E_{\nu_x}^2 e^{-E_{\nu_x}/T}$$

and an assumed detection efficiency that scales as $E_{\nu_x}^3$ (i.e. the cross section scales like $E_{\nu_x}^2$ and detection of secondary products like E_{ν_x}) gives

$$\delta t = 51.4 R_{\text{MPC}} \frac{\int E_{\nu_x}^5 \left(\frac{M_{\nu_x}}{E_{\nu_x}}\right)^2 e^{-E_{\nu_x}/T}\, dE_{\nu_x}}{\int E_{\nu_x}^2 e^{E_{\nu_x}/T}\, dE_\nu} \tag{9}$$

and for the case of μ, τ neutrinos we expect [5] $T \simeq 25/3$, giving

$$\delta t = 0.037\, M_{\nu_x}^2\, R_{\text{MPC}}\ \text{sec} \tag{10}$$

For a galactic supernova $R_{\text{MPC}} = 0.01\text{MPC}$ and for $M_{\nu_x} = 30$ eV we find

$$\delta t = 330\text{ms} \tag{11}$$

Note that the mean time separation and shape of the time distribution are altered in a characteristic manner by the different neutrino masses. It is this characteristic that must be used to extract a cosmologically significant mass from a Galactic Supernova.

There are two possibilities for the detection of a finite neutrino mass – both illustrated in Figure 4.[8]

(i) Detection of a large time difference using an extra galactic supernova (Figure 4b)

(ii) Detection of a small time difference using a galactic supernova (Figure 4b)

The detailed time distribution for a galactic supernova is shown in Figure 5. Table 1 lists the reactions that can be used for supernova burst detection and the experimental techniques that can be employed. Note from Table 1 that no detector will provide information on all the possible channels. In this note we are mainly concerned with the prospects for extracting a neutrino mass of the μ and τ neutrino if the mass value is in the 10 - 40 eV mass range of cosmological significance. Note also that the mean time distribution for these cases are

$$\delta t = 35 \text{ ms} \qquad M_{\nu_x} = 10 \text{ eV}$$
$$\delta t = 590 \text{ ms} \qquad M_{\nu_x} = 40 \text{ eV}$$

In a sense the galaxy is simply too small to obtain large time differences for cosmologically interesting neutrino mass. From Fig. 1 it is clear that the shape of the time pulse changes with M_{ν_x} and that the mean width of the initial pulse is

$$\delta t \simeq 400 \text{ ms} \quad \text{for} \quad M_{\nu_x} \simeq 10 \text{ eV}$$
$$\delta t \simeq 1 \text{ sec} \quad \text{for} \quad M_{\nu_x} \simeq 40 \text{ eV}$$

and thus

$$(\delta t)_{\text{pulse width}} > (\delta t)_{\text{time difference } (-) \text{ between } \overset{(-)}{\nu}_x \text{ and } \overset{(-)}{\nu}_x \text{ neutrino arrival}}$$

It is clear that a very large number of events and very good time resolution is required to resolve the effects in the lower neutrino mass range due to the shape of the time pulse near the origin. We propose to use the derivative of the pulse to give a zero crossing estimate of the arrival times to obtain the required accuracy.

We now consider the expected event ratio for various channels for planned or proposed detectors for supernova detection in the 1990's and beyond. Table 2 lists the approximate event rates for several detectors in the construction or planning stage (ICARUS, SNO, LVD, MACRO) and for two newly proposed detectors:[4,8,9,10,11]

i) Super Kamiokande[11]

ii) SNBO (Supernova Neutrino Burst Observatory) as well as the existing IMB and Kamio-kande II detectors [other detectors are likely too small to give additional information.]

We refer to Ref. [11] for discussions of the proposed Super Kamiokande detector. The SNBO detector would have the active mass of 100,000 tons of $CaCO_3$ and would be instrumented with a large number of neutron detectors. The detector concept has been described in Refs. [5] and [8]. This detector is mainly sensitive to ν_μ and ν_τ neutrinos due to the dynamics

of the neutral current process which strongly discriminates against lower energy ν_e and $\bar{\nu}_e$ events, thus, it is a ν_μ, ν_τ detector. Therefore, this zero crossing technique could be applied to obtain the desired time resolution.

From Table 2 it is clear that most planned detectors will not have the ability to uniquely detect a (10 – 40) eV neutrino mass signal. This is primarily due to the fact that

(i) The $\nu_{\mu,\tau}e \to \nu_{\mu,\tau}e$ signal rate is very small and much less than the $\nu_e e \to \nu_e e$ rate.

(ii) For the $\nu_{\mu,\tau}d \to (np)\nu_{\mu,\tau}$ rate in SNO only the time of arrival of the neutrino is measured and the uncertainly in the timing structure shown in Fig 1 and the background from $\overset{(-)}{\nu}_e$ events may make it difficult to uniquely extract the ν_μ or ν_τ signature

Note that the SNBO detector could also provide a higher mass limit on one of the neutrinos if the rate of events is considerably below the prediction. This would imply that one of the neutrinos has a mass of > 100 eV or is unstable. The two newly proposed detectors provide additional information that could be used to uniquely detect a ν_μ or ν_τ mass even in the 10 eV range. This is due to the fact that the Super Kamiokande detector gives adequate numbers of $\nu_x e \to \nu_x e$ events to possibly make the separation in the high mass case and the SNBO detector (when combined with the other detector results) provides a very large number of pure ν_μ and ν_τ events. With 5000 ν_μ and 5000 ν_τ events it should be possible to make a unique separation of low mass neutrinos (i.e. $M_{\nu_\mu} \simeq 0$; $M_{\nu_\tau} \sim 30$ eV) and mixed cases such as ($M_{\nu_\mu} \simeq 10$ eV, $M_{\nu_\tau} \sim 30$ eV). [Note that the zero crossing technique is difficult to use in the case of a mixed, overlapping ν_e, ν_x sample.]

The construction and long term operation of such detectors in a self triggered "Supernova Watch" detection mode is essential to determine if the μ and τ (and possibly ν_e) neutrinos have mass values that are important for the cosmology of the universe. At this time there appears to be no other viable proposal of techniques to carry out the important measurement. In addition to the neutrino mass determination the detection of thousands of $\bar{\nu}_e$, ν_e and ν_μ/ν_τ events from a future supernova will provide crucial information about other properties of neutrinos (such as a magnetic moment in the range of $10^{-11} - 10^{-14}\mu_B$) and exotic $\nu\nu$ interactions as well as the dynamics of stellar collapse and explosion.

The SNBO could be constructed in a limestone deposit in places like Arizona. Figure 6 shows a conceptual design of such a detector. Prototype neutron detectors are being developed by the collaboration (see Figure 7).

Now we can compare the prospects for detecting a finite mass neutrino from the use of galactic and extra galactic supernova as given in Table 3. While the lowest mass detector is

one using very low temperature sensors (such as 300 mK superconducting grams) the most easily constructed is the SNBO.[12]

There is another technique to observe a massive ν or τ neutrino through the neutrino oscillation experiment

$$\nu_\mu \to \nu_\tau; \ \nu_\tau + N \to \tau + X$$

$$\text{mass } \nu_\mu \text{ or } \nu_\tau \sim 30\,\text{eV}$$

Two techniques have been proposed for this process using a high energy neutrino beam at FNAL or CERN

(i) ~ 1 ton of Nuclear Emulsion

(ii) A Several Ton Liquid Argon or Krypton Detector

(by UCLA and also the ICARUS Groups)

We illustrate the (UCLA, ICARUS) method in Table 4. Mixing angles as low as $\sin^2 2\alpha \sim 10^{-4}$.

In summary, Neutrino Dark Matter is still possible and two techniques to search for it were reported here.

REFERENCES

[1] Li Zhi Fang, Shou-Ping Xang and Lin Yan, Biased Clustering in an Universe with Hot Dark Matter and Cosmic Strings, Fermilab Pub 90/23-A, 1990.

[2] See for example E.W.Kolb, D.N.Schramm and M.Turner, in Neutrino Physics, (ed. K.Winter, Cambridge Univ.Press, 1989).

[3] R.Cowsik and J.McClelland, Phys.Rev.Lett., 29, 660 (1972).

[4] See for example, A.Burrows, report in Observational Neutrino Astronomy, (ed D.Cline, World Scientific, 1988).

[5] D.Cline, et.al., A New Method for Detection of Distant Supernova Neutrino Bursts, to be published in Astro.Phys.Letts and Comm. (1989).

[6] J.Wilson and R.Mayle, Livermore Preprint (1988) and to appear in Ap.J.

[7] T.Piran and J.Wilson (unpublished) and D.Cline, et.al, Proposal to Study a New Type of Neutrino Burst Detector: The Supernova Neutrino Burst Observatory (SNBO), UCLA preprint 1990.

[8] D.Cline, Neutrino Astronomy, the proceeding of the 14th Texas Symposium on Relativistic Astro Physics, Dallas, Texas, (1988), to be published by the New York Academy of Sciences.

[9] A.Burrows, M.Turner and R.E.Brickman, Phys.Rev, D (1989).

[10] The event rate estimates for the SNO detector come from P.Doe, "The SNO Detector", published in Observational Neutrino Astronomy, (ed. D.Cline, World Scientific, 1988, p.92).

[11] The Super Kamiokande proposal, Univ. of Tokyo, (1988) and M. Koshiba, private communication.

[12] Private Communication, P.Smith

TABLE 1

SUPERNOVA NEUTRINO DETECTION IN THE 1990'S				
REACTIONS:	$\bar{\nu}_e p \to e^+ n$	$\nu_x e \to \nu_x e$	$\nu_x N \to \nu_x N$	$\nu_x N \to \nu_x N^*$ $\hookrightarrow n$
PARAMETERS:				
CROSS SECTION	LARGE (KII, SK, IMB, LVD)	SMALL $\sim E_\nu^2$ (ICARUS)	LARGE For Coherent Process	LARGE At High E_{ν_x} SNO/SNBO
NEUTRINO ENERGY ESTIMATE	YES $\sim E_e$	PARTIAL $E_{\nu_e} \sim f(E_e)$	NO	NO But a threshold may set E_{ν_x}
ν DIRECTION	NO	YES	NO	NO
TIME INFORMATION	YES	YES	YES	YES
DOWN TIME (guess)	$\gtrsim 10\%$	$\sim 30\%$	?	(could be small)
MAXIMUM DETECTOR SIZE	2×10^5 Tons (H_2O) LENA $\lesssim 10^4$ Tons Liq. Scint. (LVD)	$\sim 2 \times 10^5$ Tons (H_2O) or $\sim 10^3$ Tons Cryogenic (ICARUS)	? Kilograms No detector proposed so far	$\sim 10^5 - 10^7$ Tons of $CaCO_2$ (SNBO) or $\sim 10^3$ Tons D_2O(SNO)
BACKGROUNDS	SMALL If e^+ and n capture detected - OK for H_2O Galactic Signal	SMALL If directionality used to reject background	?	DEPENDS on Radioactivity of material

TABLE 2

COMPARISION OF FUTURE SUPERNOVA ν DETECTORS						
PROCESS:	$\bar{\nu}_e p \to e^+ n$ $\bar{\nu}_e d \to ppe^-$	$\bar{\nu}_e e \to \bar{\nu}_e e$	$\bar{\nu}_x e \to \bar{\nu}_x e$ $x = \mu, \tau$	$\nu_e N \to N^* \nu_e$ $\hookrightarrow$ n	$\nu_\mu N \to N^* \nu_x$ $\hookrightarrow$ n	ν_e Prompt
<u>DETECTORS:</u>						
ICARUS (3kT)	–	~ 140	25	–	–	4*
SNO (1kT)	~ 500	60	20	~ 200	~ 400	5*
(D$_2$O + H$_2$O)	(H$_2$O Shield + D$_2$O)			–	–	5 – 20*
LVD/MACRO (3kT) scent	~ 1000	–	–	–	–	
Kam.II/IMB	(~ 480)	(~ 60)	(~ 20)	–	–	–
SUPER Kam. (30 kT) H$_2$O	~ 4000	~ 600	200	–	–	~ 5*
SNBO (100kT)	(100's)?			~ 100's	10,000	–
COMMENTS:	measure t_ν, $E_\nu \sim E_e$ No Direction	$t\nu$ E_ν estimated from E_e Θ_ν Measured		$t\nu$ only No E_ν ! No Θ_ν		$\Delta t \simeq 10$ ms

* Depends on Energy Spectrum of Prompt ν_e and Detector Threshold

TABLE 3

		Mass	Rate (ν_x)	Distance	M_ν
To Measure $M_\nu \sim 2\,eV$ Must Use Extragalactic Supernova					
H_2O	3×10^4 T (Super Kam.)	~ 200	10 Kpc	20 eV
	$(10^8$ T$)$	~ 200	0.6 Mpc	2 eV
$CaCO_2$	10^7 T (SNBO)	2×10^4	10 Kpc	20 eV
	(10^7) T	300	0.6 Mpc	2 eV
Cryogenic	10 T*	30*	10 Kpc*	20 eV*
$(T < 300\,MK)$	(10^4) T*	30*	0.6 Mpc*	2 eV*

* $\Delta t \sim 150\,ms$

P. Smith, private communication

TABLE 4

A Possible Technique to Detect $\nu_\mu \to \nu_\tau$

at the $\mathrm{Sin}^2 2\alpha \sim 10^{-4}$ Level Using a

Liquid Argon Tracking Chamber

It is possible that $m_{\nu_\tau} \sim 30\,\mathrm{eV}$ and $m_{\nu_\mu} \sim 0\,\mathrm{eV}$ such that $\delta^2 m \sim 900\ \mathrm{eV}^2$. In this case we would like to search for $\nu_\mu \to \nu_\tau$ in a high energy neutrino beam at FNAL or CERN. We propose the following technique using a multiton ICARUS-like tracking chamber.

(i) $\nu_\tau + \mathrm{N} \to \tau + \mathrm{X}$.
   Possible rejection factor $N_{\nu_e}/N_{\nu_\mu} \sim 10^{-2}$

(ii) Require missing transverse energy due to $\tau \to e\nu\bar{\nu}$.
   Possible refection factor $\sim 10^{-1}$

(iii) Observe effects of τ lifetime. This requires high resolution at the vertex. The ICARUS tests have achieved $\sim 60\,\mu\mathrm{m}$ in the Z direction.
   Possible reduction factor $\sim 10^{-1}$.

For a 10 ton detector and 10^{19} protons on target we collect $> 10^5\,\nu_\mu$ interactions. For $\mathrm{Sin}^2 2\alpha \sim 5\times10^{-4}$ we sould produce $\sim (\,10\ \mathrm{to}\ 20\,)\,\nu_\tau + \mathrm{N} \to \tau + \mathrm{X}$ events. Additional calorimeters and μ tracking would be required ar•und the detector in order to establish the effects of the cuts (i), (ii) and (iii).

EXAMPLE – FNAL Pub. –   Li Zhi Fang, S.P. Xieng, Len Yen

"Biased Clustering in a Universe with
Hot Dark Matter and Cosmic Strings"

Model – Cosmic Strings Seed Galaxy Baryons and Neutrinos form Halos –

The size of Galactic clusters is related to the Neutrino Halo

$M_\nu \sim (10 - 100)\,\text{eV}$

V (Km/s) vs r (Kpc)

Figure 1

Wilson

20

$\bar{\epsilon}_{\nu\mu,\tau}$ (Escape from Core!)

15

$\bar{\epsilon}_{\bar{\nu}_e}$

MeV

$\bar{\epsilon}_{\nu_e}$

10

$L_{\bar{\nu}_e}$

$\longrightarrow$ Explosion Starts

5

Cooling By $\nu\bar{\nu}$

Rise of

Accretion Pulse

0   0 5   1.0   1.5

Time sec

$\longleftrightarrow$

Accretion Pulse

Figure 2

Mass Measurement of Time of Flight Difference

Accelerator
δt

Measure
Δt

P   Target M_2 Detector

L

$$\Delta t \propto L \cdot \left[\left(\frac{M_1}{E_1}\right)^2 - \left(\frac{M_2}{E_2}\right)^2 \right]$$

Normally $\Delta t \gg \delta t \gg Detector$

$TimeResolution$

Supernova

ν_1, ν_2

Earth

R

$\delta t \sim \Delta t$

Intrinsic Time Spread δt Larger Than Time Of Flight Difference

Figure 3

Case i $\vdash$ Δt $\dashv$ for $M_\nu \sim 30\,eV$

Counts

100

$\bar{\nu}_e$ ν_μ or ν_τ

10

?

ν_e $\rightarrow$ $\leftarrow$ M_ν

$\sim 10\,eV$

1

Extragalactic
Source

1 ms   10   100 ms   1 sec   10 sec   100 sec

t_ν

Case ii

Counts

100

$\bar{\nu}_e$ ν_μ / ν_τ

$M_\nu > 10\,eV$

10

?

ν_e $\rightarrow$ $\leftarrow$ $\delta t \sim 340\,ms$

$m_\nu = 30\,eV$

1

Galactic
Supernova

1 ms   10   100 ms   1 sec   10 sec   100 sec

t_ν

Figure 4

Figure 5. Time distribution for the neutrino interactions from a Galactic Supernova as a function of the neutrino mass. The solid curves refer to ν_e and the dashed to ν_μ / ν_τ.

Conceptual Detector Location

(10^5 Tons)

$30\,\text{m} \times 30\,\text{m} \times 100\,\text{m}$
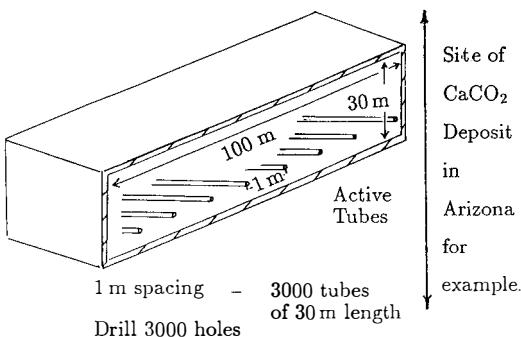
Figure 6
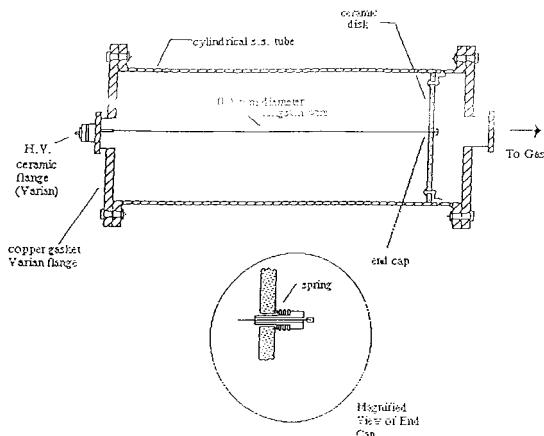
A Schematic of prototype BF$_3$ detector / UCLA
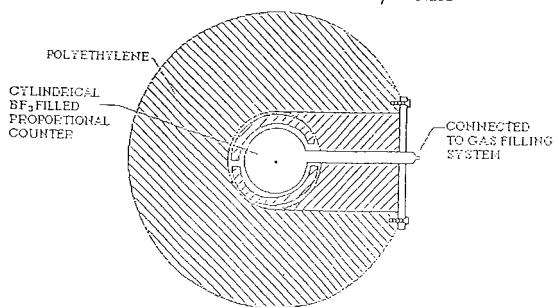
Figure 7

NEUTRINO RESONANT OSCILLATION AND SPIN-FLAVOUR PRECESSION

AND

NEW $\mathcal{Y}$-SIGNAL FROM SUPERNOVAE.

E. Kh. AKHMEDOV[a] and Z.G. BEREZHIANI[b]

[a]*Kurchatov Institute of Atomic Energy, Moscow 123182, USSR*

[b]*Institute of Physics of the Academy of Sciences of*
Georgian SSR, Tbilisi 380077, USSR

ABSTRACT
We study resonant neutrino oscillations (RNO) and resonant spin-flavor precession (RSFP) of Majorana neutrinos in supernovae. The response of water and heavy water detectors to supernova neutrinos is calculated for the following scenarios: (i) Standard supernova model; (ii) RNO; (iii) RSFP, and (iv) RNO+RSFP. For the water detectors, expected event numbers are calculated using the Kamiokande II detection efficiency. It is shown that the combined analysis of the isotropic/directional event-number ratio and prompt neutronization burst signal will enable one to distinguish between all the considered possibilities provided the distance to supernova L is <10 kpc. Even more rich and conclusive information may be reached by using the D_2O detector.

Presented by Z.G. Berezhiani.

1. If neutrino (ν) oscillations do occur, matter can enhance them significantly, leading to resonant transitions between various ν flavors[1,2] Recently, it has been shown that resonant conversions can also take place if the ν's have transition magnetic moments μ_{ij}[3-6] In this case, in the presence of a transverse magnetic field, ν spin rotates with its flavor changing simultaneously. Matter can amplify resonantly such a spin-flavor precession, much in the same manner as it enhances the ν oscillations. Both the resonant ν oscillations (RNO) and resonant spin-flavor precession (RSFP) may account for the observed solar-ν deficiency.

Neutrino flavor conversions due to RNO and RSFP may have important consequences also for ν's from supernovae (SN). Implications of the RNO for SN ν's were studied in a number of papers (see, e.g.. refs. 7-9 and references therein). As concerns to RSFP, only a few qualitative comments on its consequences for ν's from SN were made[4,6] The main goal of the present report is to study these consequences in some detail and to compare them with those of the RNO and with the predictions of the standard SN model.

We study here only the RSFP of Majorana ν's; implications of the Dirac ν transition magnetic moments for SN dynamics and the ν signal from SN will be considered elsewhere.

2. We assume that all the ν mixing angles are small and there is a direct mass hierarchy $m_{\nu_e} < m_{\nu_\mu} < m_{\nu_\tau}$. Hence the RNO may occur for ν's and not for antineutrinos. It is only in this case that the RNO can account for the solar-ν deficiency. We also suppose that all the ν masses are $\leqslant 3 \cdot 10^2$ eV, so that all the RNO and RSFP ν resonant conversions occur outside the ν-spheres.

The ν conversions due to RNO and RSFP will be efficient only if the corresponding adiabaticity conditions are satisfied, i.e. if the oscillation or precession length at resonance is much smaller than the size of the resonant enhancement region[2-6] In what follows we shall consider four scenarios: (i) There are no resonant ν conversions (standard model, SM), (ii) All the possible RNO conversions are adiabatic whereas the RSFP ones are not (RNO), (iii) All the RSFP conversions are adiabatic whereas the RNO ones are not (RSFP), and (iv) All the RNO and RSFP conversions are adiabatic (RNO+RSFP).

As SN ν's move outwards from the ν-spheres, they pass through several RNO and RSFP resonances. The order of these resonances and

their character depend crucially on the magnitude of the electron number per nucleon Y_e[6]. The calculations show that $Y_e > 1/3$ outside the ν-spheres; it increases slowly with increasing distance from the center of SN[10]. The order of the resonant transitions for $1/3 < Y_e < 1/2$ is shown in Fig. 1. It should be noted that at very low densities ($\rho < 10^{-2} g/cm^3$) Y_e becomes $> 1/2$. However, at such small densities either the RNO or the RSFP or both may become non-adiabatic; in this case the corresponding ν conversion is not efficient and can be disregarded.

The fate of the ν's in SN in each scenario can be easily read off the scheme of the resonances given in Fig. 1. For the (RNO+RSFP) scenario it is displayed in Fig. 2. For the RNO case the following transmutations will result: $\nu_e \rightarrow \nu_\tau$, $\bar{\nu}_e \rightarrow \bar{\nu}_e$, $\nu_\mu \rightarrow \nu_e$, $\bar{\nu}_\mu \rightarrow \bar{\nu}_\mu$, $\nu_\tau \rightarrow \nu_e \rightarrow \nu_\mu$, $\bar{\nu}_\tau \rightarrow \bar{\nu}_\tau$. For the RSFP we shall have $\nu_e \rightarrow \nu_e$, $\bar{\nu}_e \rightarrow \nu_\tau \rightarrow \bar{\nu}_\mu$, $\nu_\mu \rightarrow \bar{\nu}_e$, $\bar{\nu}_\mu \rightarrow \nu_\tau$, $\nu_\tau \rightarrow \bar{\nu}_e \rightarrow \nu_\mu$, $\bar{\nu}_\tau \rightarrow \bar{\nu}_\tau$.
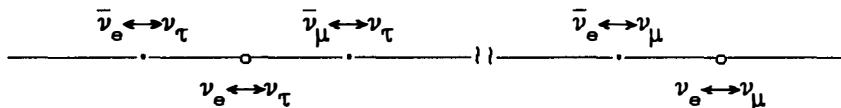
Fig. 1. Scheme of the RSFP and RNO resonances in SN. The precession resonances are indicated above the axis, and those of ν oscillations, below it. Matter density decreases from the left to the right.
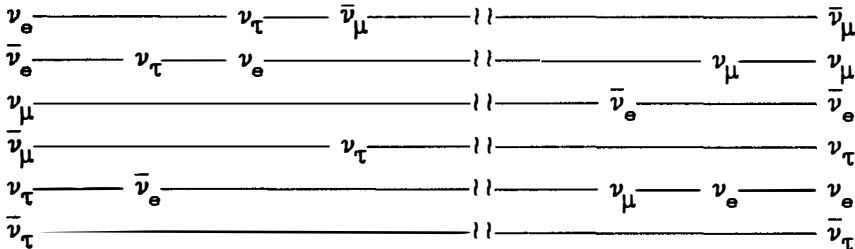
Fig. 2. Neutrino transmutations in SN (RNO+RSFP scenario).

4. In our calculations we shall assume, following refs. 10,11, that SN ν's have Fermi-Dirac spectra with zero chemical potentials, their temperatures and luminosities being

$$T_{\nu_e} = (3-6) \text{ Mev}, \quad T_{\bar{\nu}_e} = 1.1 T_{\nu_e}, \quad T_{\nu_\mu} = T_{\bar{\nu}_\mu} = T_{\nu_\tau} = T_{\bar{\nu}_\tau} = \mathscr{Z} T_{\nu_e} \tag{1}$$

$$I_{\bar{\nu}_e} = I_{\nu_e}, \quad I_{\nu_\mu} = I_{\bar{\nu}_\mu} = I_{\nu_\tau} = I_{\bar{\nu}_\tau} = k I_{\nu_e}, \quad \mathscr{Z} = 2\text{-}2.5, \quad k = 1.0\text{-}1.5$$

The spectrum of the prompt neutronization ν_e's is poorly known; we quite arbitrarily choose it to be the same as the spectrum of thermal ν_e's with the same mean energy. The SN energy loss due to the neutronization ν_e's is taken to be $\approx 3\%$ of the total energy E_0 carried away by the ν's. For Type II SN, it is expected that $E_0 = (2-4)\cdot 10^{53}$ erg[10,11]; in our calculations we take $E_0 = 3\cdot 10^{53}$ erg.

Main ν-induced reactions in H_2O detectors are the $\bar{\nu}_e$-p capture $\bar{\nu}_e p \rightarrow ne^+$ and $^{(-)}_{\nu_i}e$ scattering; in the latter process ν's of all the species can participate. These reactions can be discriminated experimentally by the angular distributions of the $e^{\pm}$ particles: For the $\bar{\nu}$-p capture, the emitted positrons are isotropic, whereas the recoil electrons from the $^{(-)}_{\nu_i}e$ scattering are forward peaked.

For the D_2O detector, in addition to the $^{(-)}_{\nu_i}e$ scattering, reactions on deuterons are also possible. They are charged current $^{(-)}_{\nu}$-d capture and neutral current $^{(-)}_{\nu}$d disintegration:

$$\nu_e d \rightarrow ppe^- \ (CC1), \quad \bar{\nu}_e d \rightarrow nne^+ \ (CC2), \quad ^{(-)}_{\nu_i}d \rightarrow np\, ^{(-)}_{\nu_i} \ (NC) \qquad (2)$$

All these reactions can be distinguished experimentally[12]. The expressions for the event numbers in H_2O and D_2O detectors as well as the relevant cross sections can be found in ref. 13.

5. Our results for the response of H_2O detectors to SN ν's are given in Tables 1-3. In Table 1, the second column gives the expected event number for $\bar{\nu}_e$-p capture, whereas the columns 3 through 8 give event numbers for $^{(-)}_{\nu_i}e$ scattering. Although it is only the sum of all the $^{(-)}_{\nu_i}e$ event numbers that is measured, we show these numbers separately to illustrate their changes due to various resonant conversions. Also shown are the prompt neutronization burst (PNB) contribution, isotropic-to-directional event ratio, $I/D \equiv (\bar{\nu}_e p)/[\Sigma_i (^{(-)}_{\nu_i}e)+PNB]$, and mean energies of recoil electrons (positrons) for the isotropic and directional events, $E^{(e)}_{isotr.}$ and $E^{(e)}_{dir.}$. In Tables 2 and 3 only the sum of all the $^{(-)}_{\nu_i}e$ event numbers is given.

Table 1.

Scenario	$\bar{\nu}_e p$	$\nu_e e$	$\bar{\nu}_e e$	$\nu_\mu e$	$\bar{\nu}_\mu e$	$\nu_\tau e$	$\bar{\nu}_\tau e$	PNB	I/D	$E^{(e)}_{isotr.}$	$E^{(e)}_{dir.}$
St. mod.	140	1.8	.39	.60	.48	.60	.48	.32	30	18	15
RNO	140	3.9	.39	.60	.48	.26	.48	$4.7\cdot 10^{-2}$	22	18	18
RSFP	330	1.8	.97	.60	.24	.60	.48	.32	66	30	15
RNO+RSFP	330	3.9	.97	.31	.20	.60	.48	$3.6\cdot 10^{-2}$	50	30	18

It can be seen from Tables 1-3 that the PNB signal is strongly reduced whenever the RNO take place. This is because in the RNO and RNO+RSFP scenarios the original ν_e's are transformed into ν_τ's or $\bar{\nu}_\mu$'s, whose νe scattering cross sections are ~7 times smaller than that of ν_e's. Therefore the suppression of the PNB signal can be considered as an indication of the RNO occurrence in SN[7-9].

Table 2.

Scenario	$\bar{\nu}_e p$	$\Sigma^{(\bar{\nu})}_i e$	PNB	I/D	$E^{(e)}_{isotr.}$	$E^{(e)}_{dir.}$	T_{ν_e} (MeV)
St. mod.	220	5.9	0.48	34	22	17	
RNO	220	7.7	$7.2 \cdot 10^{-2}$	28	22	21	4.0
RSFP	460	6.3	0.48	67	40	18	
RNO+RSFP	460	8.1	$5.6 \cdot 10^{-2}$	57	40	21	
St. mod.	290	7.2	0.61	38	28	20	
RNO	290	8.7	$9.2 \cdot 10^{-2}$	33	28	24	5.0
RSFP	590	7.6	0.61	71	50	20	
RNO+RSFP	590	9.2	$7.2 \cdot 10^{-2}$	64	50	24	
St. mod.	330	7.7	0.66	39	30	21	
RNO	330	9.1	0.10	36	30	26	5.45
RSFP	640	8.1	0.66	74	54	21	
RNO+RSFP	640	9.6	$7.9 \cdot 10^{-2}$	67	54	26	

The number of the isotropic $\bar{\nu}_e p$ events remains unaltered in the RNO scenario as compared to the SM predictions but is sizably increased whenever the RSFP takes place. This is because in the RSFP and RNO+RSFP scenarios the detected $\bar{\nu}_e$'s are the original ν_μ's, whose temperature is about twice as high as that of the original $\bar{\nu}_e$'s and, in addition, whose luminosity may be slightly higher.

All the event numbers depend on the temperatures of the ν Fermi-Dirac distributions. Therefore the increase of $(\bar{\nu}_e p)$ due to RSFP might be simulated by the suitable increase of the $\bar{\nu}_e$ temperature, which is known only to an accuracy of about a factor of two (compare the event numbers for the ν_e temperatures of 3.0 and 5.45 MeV in Tables 1 and 2). However, the distinction can be made by examining the I/D and PNB magnitudes. In Table 3 we give these values for the Kamiokande II (K II) detection efficiency. To assess possible uncertainties due to ambiguities of the SN models, we

calculated the I/D and PNB magnitudes for $k=1$, 1.5 and $æ=2$, 2.5.

Table 3.

			St.mod.	RNO	RSFP	RNO+RSP	T_{ν_e} (MeV)
	PNB		0.32	$4.7 \cdot 10^{-2}$	0.32	$3.6 \cdot 10^{-2}$	
	$k=1$,	$æ=2$	30	22	66	50	
	$k=1.5$,	$æ=2$	24	15	77	52	3.0
	$k=1$,	$æ=2.5$	28	19	78	56	
I/D	$k=1.5$,	$æ=2.5$	22	13	89	58	
	$k=1$,	$æ=2$	39	36	74	67	
	$k=1.5$,	$æ=2$	33	25	88	69	
	$k=1$,	$æ=2.5$	38	33	89	79	5.45
	$k=1.5$,	$æ=2.5$	31	24	110	81	
	PNB		0.66	0.10	0.66	$7.9 \cdot 10^{-2}$	

As it is seen from Table 3, I/D and PNB values vary with vary-
ing T_{ν_e}, $æ$ and k. However, all the I/D values can be divided into
two groups: First group, the SM and RNO scenarios, and second, the
RSFP and RNO+RSFP scenarios. Some figures inside the first group
may be confused with each other; the same is also true for the
figures inside the second group. However, no one I/D value from the
first group can be mistaken for the value from the second group
and vice versa. Therefore we can distinguish between these two
groups by measuring the I/D event ratio.

Now the question is how to discriminate between different pos-
sibilities inside each group, i.e. between the SM and RNO, or be-
tween the RSFP and RNO+RSFP scenarios. From Table 3 it is seen
that we can do that by counting the ν events from PNB. Thus we can
distinguish between all the four possibilities by studying the I/D
and PNB magnitudes.

6. Our results for the response of D_2O detector to SN ν's are
given in Tables 4 and 5. It can be seen from Table 4 that, given
the ν_e temperature, the NC signal is the same for all the four
considered scenarios. This applies not only to the thermal ν's but
to the PNB ν's as well. The invariance of the NC signals stems from
the fact that the cross sections of the NC reaction $^{(\overline{\nu})}_1 d \to np\,^{(\overline{\nu})}_1$ are
the same for all the ν species provided $E_\nu \lesssim 50$ MeV. Thus the NC-mode
ν signals can serve as very useful benchmarks.

Table 4.

Scenario	$(\bar{\nu})d\to np\,(\bar{\nu})$ (NC)		$\nu_e d\to ppe^-$ (CC1)		$\bar{\nu}_e d\to nne^+$	$\dfrac{CC1}{CC2}$	$\dfrac{NC}{CC1}$	T_{ν_e}
	Th.phase	PNB	Th.phase	PNB	(CC2)			(MeV)
St.mod.	380	5.4	65	12	51	1.3	5.9	
RNO	380	5.4	174	0	51	3.5	2.2	3.0
RSFP	380	5.4	65	12	130	0.5	5.9	
RNO+RSFP	380	5.4	174	0	130	1.3	2.2	
St.mod.	820	13	150	28	130	1.2	5.4	
RNO	820	13	390	0	130	2.9	2.1	5.45
RSFP	820	13	150	28	300	0.5	5.4	
RNO+RSFP	820	13	390	0	300	1.3	2.1	

All the changes of the event numbers due to various resonant ν conversions can be easily understood considering the ν transmutations in each scenario.

Table 5.

	CC1/CC2				NC/CC1				T_{ν_e}
	SM	RNO	RSFP	RNO+RSFP	SM	RNO	RSFP	RNO+RSFP	(MeV)
$k=1$, $x=2$	1.3	3.5	0.49	1.3	5.9	2.2	5.9	2.2	
$k=1.5$, $x=2$	1.3	5.2	0.33	1.3	8.3	2.1	8.3	2.1	3.0
$k=1$, $x=2.5$	1.3	4.7	0.36	1.3	7.5	2.1	7.5	2.1	
$k=1.5$, $x=2.5$	1.3	7.0	0.24	1.3	11	2.0	11	2.0	
$k=1$, $x=2$	1.2	2.9	0.51	1.3	5.4	2.1	5.4	2.1	
$k=1.5$, $x=2$	1.2	4.3	0.34	1.3	7.5	2.0	7.5	2.0	5.45
$k=1$, $x=2.5$	1.2	3.9	0.38	1.3	6.7	2.0	6.7	2.0	
$k=1.5$, $x=2.5$	1.2	5.8	0.26	1.3	9.5	1.9	9.5	1.9	

In Table 5 the event-number ratios CC1/CC2 and NC/CC1 are presented for two values of T_{ν_e} and various magnitudes of k and x. It can be seen from this table that for the SM and RNO+RSFP scenarios, the CC1/CC2 ratio is practically the same, whereas for the RNO (RSFP) case this ratio is noticeably larger (smaller). The NC/CC1 ratio distinguishes the SM and RSFP scenarios from the RNO and RNO+RSFP ones. Thus, by comparing the CC1/CC2 and NC/CC1 event ratios, one will be able to discriminate between all the four considered scenarios. The PNB contributions to NC and CC1 channels can then serve for an additional check of consistency (see Table 4).

312

7. In our consideration we assumed that either all the conversions of a given type are adiabatic or all of them are non-adiabatic. However, this need not be the case. We have studied the most plausible alternative possibilities. Our main conclusion is that either all the consequences coincide with those derived in Sects. 5 and 6, or there is a minor difference for the RNO+RSFP scenario: The observed ν_e's are the original $\bar{\nu}_e$'s and therefore this case is similar to the RSFP one; however, the RSFP and RNO+RSFP cases can be told by the FNB signal since it should be unchanged for the former scenario and suppressed for the latter one (see ref. 13 for more details).

All the qualitative conclusions we derived for the D_2O detector are also applicable to the ^{11}B one[14] or to any other detector allowing the comparison of the NC channel with the two CC channels.

In conclusion, we calculated the response of water and heavy water detectors to the ν signal from SN allowing for the resonant flavor and spin-flavor conversions of ν's. It is shown that for a nearby (L≤10 kpc) SN or sufficiently large detectors the analysis of the ν signal will allow one to reveal whether any resonant ν conversion occurs in SN and what its nature is. The detection of SN ν's may also shed some new light on the solar-ν problem.

The authors are grateful to S.I. Blinnikov, A. Burrows, D.K. Nadyozhin and A.Yu. Smirnov for useful discussions.

References

1. L. Wolfenstein, Phys. Rev. D17 (1978) 2369.
2. S.P. Mikheyev and A.Yu.Smirnov, Sov.J. Nucl. Phys. 42(1985)913.
3. E.Kh. Akhmedov, preprint IAE-4568/1, January 1988.
4. C.-S. Lim and W.J. Marciano, Phys. Rev.D37 (1988) 1368.
5. E.Kh. Akhmedov, Phys. Lett. B213 (1988) 64.
6. E.Kh. Akhmedov, Sov. Phys. JETP 68 (1989) 676.
7. S.P. Mikheyev and A.Yu. Smirnov, Sov. Phys. JETP 64 (1986) 1; Pisma ZhETF 46 (1987) 11.
8. T.P. Walker and D.N. Schramm, Phys. Lett. B195 (1987) 331.
9. H. Minakata and H. Nunokawa, preprint TMUP-HEL-8913, 1989.
10.S.W. Bruenn, Astrophys. J. Suppl. Ser. 58 (1985) 771.
11.A. Burrows, Astrophys. J. 334 (1988) 891.
12.G. Aardsma et al., Phys. Lett. B194 (1987) 321.
13.E.Kh. Akhmedov and Z.G. Berezhiani, preprint IAE-5058/1, 1990.
14.R.S. Raghavan, S. Pakvasa and B.A. Brown, Phys. Rev. Lett. 57 (1986) 1801; R.S. Raghavan, Phys. Rev. D34 (1986) 2088.

LEP (ASTRO) PHYSICS

COSMOLOGY AND EXPERIMENTAL PARTICLE PHYSICS

David N. Schramm

The University of Chicago,
5640 S. Ellis Avenue, Chicago, IL 60637
and
NASA/Fermilab Astrophysics Center,
Fermi National Accelerator Laboratory
Box 500, Batavia, IL 60510-0500

ABSTRACT

With LEP and SLC, particle accelerators are now able to test the cosmological standard model in the same manner in which they test the standard model of the particle world. In particular, LEP and SLC have now verified the cosmological models requirement of $N_\nu \sim 3$. Furthermore, LEP is now constraining possible non-baryonic dark matter candidates in a significant way. Thus, one can say that cosmology now uses accelerators as well as telescopes for verification of ideas.

INTRODUCTION

Traditionally, cosmological ideas were influenced by observers using telescopes, while particle physics ideas were influenced by experimenters using accelerators. However, in recent years a relationship has developed where a theoretical interface between these two fields has developed. However, this year we are beginning to see actual experimental work in accelerator labs that has dramatic implications for cosmology. In particular, the recent results from LEP and SLC make the particle/cosmology connection one with experimental as well as theoretical consequences.

In some sense, LEP and SLC have positively tested the standard model of cosmology, the Big Bang, in much the same way they have positively tested the standard model of particle physics, $SU(3) \times SU(2) \times U(1)$. LEP and SLC probe the Big Bang in two ways:

1) through nucleosynthesis and neutrino counting; and

2) through limiting dark matter candidates.

This particular discussion will focus on the first of these, but it is important to realize that the nucleosynthesis arguments are the definitive arguments for non/baryonic matter, thus by LEP and SLC supporting the standard Big Bang Nucleosynthesis results, particle experiments are also indirectly supporting the argument for non-baryonic matter which collider results do constrain. In this latter point (No. 2 above), LEP and SLC are supplemented by CDF and UA2 as well as by underground searches.

As to Big Bang Nucleosynthesis (BBN) itself, it is worth remembering that along with the 3K background radiation, the agreement of the observed light element abundances with the nucleosynthetic predictions is one of the major cornerstones of the Big Bang itself. The new COBE[1] results have given renewed confidence in the $3K$ background argument, just as LEP has given us renewed confidence in the BBN arguments. Because the microwave background probes events at temperatures $\sim 10^4 K$ and times of $\sim 10^5$ years, whereas the light element abundances probe the Universe at temperatures $\sim 10^{10} K$ and times of ~ 1 sec, it is the nucleosynthesis results that have led to the particle-cosmology merger we have seen over the last decade.

HISTORY OF BIG BANG NUCLEOSYNTHESIS

Before going into the specific argument as to sensitivity of BBN to the number of neutrino families (N_ν), let us review the history of BBN. In particular, it should be noted that there is a symbiotic connection between BBN and the $3K$ background dating back to Gamow and his associates, Alpher and Herman. The initial BBN calculations of Gamow's group[2] assumed pure neutrons as an initial condition and thus were not particularly accurate, but their inaccuracies had little effect on the group's predictions for a background radiation.

Once Hayashi (1950) recognized the role of neutron-proton equilibration, the framework for BBN calculations themselves has not varied significantly. The work of Alpher, Follin and Herman[3] and Taylor and Hoyle[4], preceeding the discovery of the 3K background, and Peebles[5] and Wagoner, Fowler and Hoyle,[6] immediately following the discovery, and the more recent work by the Chicago group and others[7,8,9,10] all do essentially the same basic calculation, the results of which are shown in Figure 1. As far as the calculation itself goes, solving the reaction network is relatively simple by the standards of explosive nucleosynthesis calculations in supernovae, with the changes over the last 25 years being mainly in terms of more recent nuclear reaction rates as input, not as any great calculational insight (although the calculational techniques have been modernized a bit[11,12].

With the exception of the effects of elementary particle assumptions to which we will return, the real excitement for BBN over the last 25 years has not really been in redoing the calculation. Instead, the true scientific action has focused on understanding the evolution of the light element abundances and using that information to make powerful conclusions. In particular, in the 1960's, the main focus was on ^4He which is very insensitive to the baryon density. The agreement between BBN predictions and observations helped support the basic Big Bang model but gave no significant information at that time with regard to density. In fact, in the mid-1960's, the other light isotopes (which are, in principle, capable of giving density information) were generally assumed to have been made during the t-tauri phase of stellar evolution,[13] and so, were
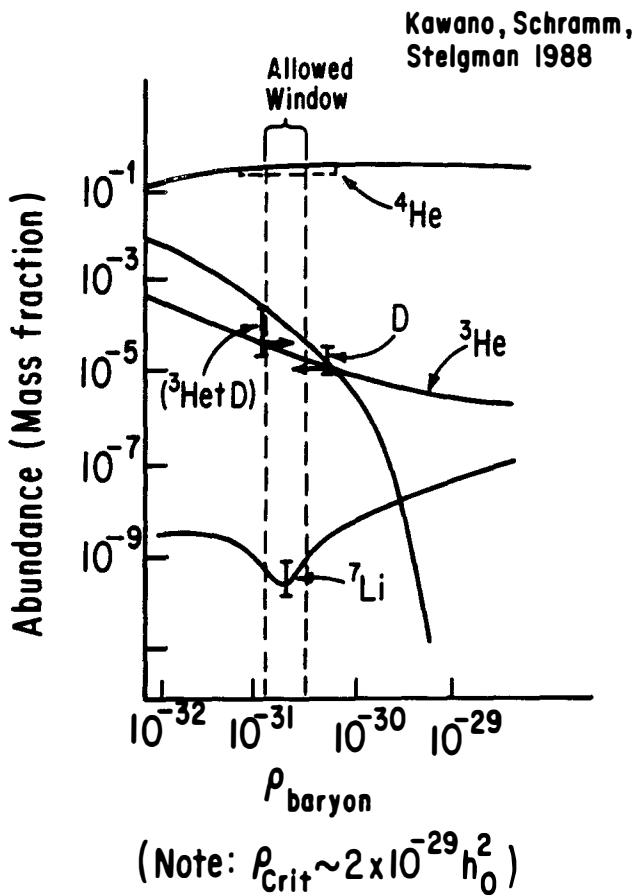
Figure 1. Big Bang Nucleosynthesis abundance yields (mass fraction) versus baryon density for a homogeneous universe.

not then taken to have cosmological significance. It was during the 1970's that BBN fully developed as a tool for probing the Universe. Here at Moriond it is particularly appropriate to note the strong role played by the Paris group in this development. Specifically, Ryter, Reeves, Gradstajn and Audouze[14] started the 1970's off by showing that the t-tauri mechanism for light element synthesis failed.

Also during the 1970's, ^2H abundance determinations improved significantly with solar wind measurements[15] and the interstellar work from the Copernicus satellite.[16] Reeves, Audouze, Fowler and Schramm[17] argued for cosmological ^2H (and ^7Li) and were able to place a constraint on the baryon density excluding a universe closed with baryons. Subsequently, the ^2H arguments were cemented when Epstein, Lattimer and Schramm[18] proved that no realistic astrophysical process other than the Big Bang could produce significant ^2H. It was also interesting that the baryon density implied by BBN was in good agreement with the density implied by the dark galactic halos.[19]

By the late 1970's, a complimentary argument to ^2H had also developed using ^3He. In particular, it was argued[20] that, unlike ^2H, ^3He was made in stars; thus, its abundance would increase with time. Since ^3He like ^2H monotonically decreased with cosmological baryon density, this argument could be used to place a lower limit on the baryon density[21] using ^3He measurements from solar wind[15] or interstellar determinations.[22] Since the bulk of the ^2H was converted in stars to ^3He, the constraint was shown to be quite restrictive.[8]

It was interesting that the lower boundary from ^3He and the upper boundary from ^2H yielded the requirement that ^7Li be near its minimum of $^7Li/H \sim 10^{-10}$, which was verified by the Pop II Li measurements of the French husband and wife team, Spite and Spite,[23] hence, yielding the situation emphasized by Yang et al.[8] that the light element abundances are consistent over nine orders of magnitude with BBN, but only if the cosmological baryon density is constrained to be around 6% of the critical value.

The other development of the 70's for BBN was the explicit calculation of Steigman, Schramm and Gunn,[24] showing that the number of neutrino generations, N_ν, had to be small to avoid overproduction of ^4He. (Earlier work had noted a dependency of the ^4He

abundance on assumptions about the fraction of the cosmological stress-energy in exotic particles,[25,4] but had not actually made an explicit calculation probing the quantity of interest to particle physicists, N_ν.) To put this in perspective, one should remember that the mid-1970's also saw the discovery of charm, bottom and tau, so that it almost seemed as if each new detection produced new particle discoveries, and yet, cosmology was arguing against this "conventional" wisdom. Over the years this cosmological limit on N_ν improved with ^4He abundance measurements, neutron lifetime measurements and with limits on the lower bound to the baryon density; hovering at $N_\nu \lesssim 4$ for most of the 1980's and dropping to slightly lower than $4^{[26,9]}$ just before LEP and SLC turned on.

BIG BANG NUCLEOSYNTHESIS: Ω_b AND N_ν

The power of Big Bang Nucleosynthesis comes from the fact that essentially all of the physics input is well determined in the terrestrial laboratory. The appropriate temperatures, 0.1 to $1MeV$, are well explored in nuclear physics labs. Thus, what nuclei do under such conditions is not a matter of guesswork, but is precisely known. In fact, it is known for these temperatures far better than it is for the centers of stars like our sun. The center of the sun is only a little over $1keV$. Thus, temperatures are below the energy where nuclear reaction rates yield significant results in laboratory experiments, and only the long times and higher densities available in stars enable anything to take place.

To calculate what happens in the Big Bang, all one has to do is follow what a gas of baryons with density ρ_b does as the universe expands and cools. As far as nuclear reactions are concerned, the only relevant region is from a little above $1MeV$ ($\sim 10^{10}K$) down to a little below $100keV$ ($\sim 10^9K$). At higher temperatures, no complex nuclei other than free single neutrons and protons can exist, and the ratio of neutrons to protrons, n/p, is just determined by $n/p = e^{-Q/T}$, where

$$Q = (m_n - m_p)c^2 \sim 1.3MeV$$

Equilibrium applies because the weak interaction rates are much faster than the expansion of the universe at temperatures much above $10^{10}K$. At temperatures much below 10^9K, the electrostatic repulsion of nuclei prevents nuclear reactions from proceeding as fast as the cosmological expansion separates the particles.

Because of the equilibrium existing for temperatures much above $10^{10}K$, we don't have to worry about what went on in the universe at higher temperatures. Thus, we can start our calculation at $10MeV$ and not worry about speculative physics like the theory of everything (T.O.E.), or grand unifying theories (GUTs), as long as a gas of neutrons and protons exists in thermal equilibriuim by the time the universe has cooled to $\sim 10MeV$.

After the weak interaction drops out of equilibrium, a little above $10^{10}K$, the ratio of neutrons to protons changes more slowly due to free neutrons decaying to protons, and similar transformations of neutrons to protons via interactions with the ambient leptons. By the time the universe reaches 10^9K $(0.1MeV)$, the ratio is slightly below $1/7$. For temperatures above 10^9K, no significant abundance of complex nuclei can exist due to the continued existence of gammas with greater than MeV energies. Note that the high photon to baryon ratio in the universe $(\sim 10^{10})$ enables significant population of the MeV high energy Boltzman tail until $T \lesssim 0.1\ MeV$. Once the temperature drops to about 10^9K, nuclei can exist in statistical equilibrium through reactions such as $n + p \leftrightarrow^2 H + \gamma$ and $H + p \leftrightarrow^3 He + \gamma$ and $^2D + n \leftrightarrow^3 H + \gamma$, which in turn react to yield ^4He. Since ^4He is the most tightly bound nucleus in the region, the flow of reactions converts almost all the neutrons that exist at 10^9K into ^4He. The flow essentially stops there because there are no stable nuclei at either mass-5 or mass-8. Since the baryon density at Big Bang Nucleosynthesis is relatively low (much less than $1g/cm^3$), only reactions involving two-particle collisions occur. It can be seen that combining the most abundant nuclei, protons, and ^4He via two body interactions always leads to unstable mass-5. Even when one combines ^4He with rarer nuclei like ^3H or ^3He, we still get only to mass-7, which, when hit by a proton, the most abundant nucleus around, yields mass-8. (A loophole around the mass-8 gap can be found if

$n/p > 1$ so that excess neutrons exist, but for the standard case $n/p < 1$). Eventually, ^3H radioactively decays to ^3He, and any mass-7 made radioactively decays to ^7Li. Thus, Big Bang Nucleosynthesis makes ^4He with traces of ^2H, ^3He, and ^7Li. (Also, all the protons left over that did not capture neutrons remain as hydrogen.) For standard homogeneous BBN, all other chemical elements are made later in stars and in related processes. (Stars jump the mass-5 and -8 instability by having gravity compress the matter to sufficient densities and have much longer times available so that three-body collisions can occur.) With the possible exception of ^7Li,[8,11,27] the results are rather insensitive to the detailed nuclear reaction rates. This insensitivity was discussed in ref. [8] and most recently using a Monte Carlo study by Krauss and Romanelli[27] An n/p ratio of $\sim 1/7$ yields a ^4He primordial mass fraction,

$$Y_p = \frac{2n/p}{n/p + 1} \approx \frac{1}{4}$$

The only parameter we can easily vary in such calculations is the density that corresponds to a given temperature. From the thermodynamics of an expanding universe we know that $\rho_b \propto T^3$; thus, we can relate the baryon density at $10^{11} K$ to the baryon density today, when the temperature is about 3 K. The problem is that we don't know today's ρ_b, so the calculation is carried out for a range in ρ_b. Another aspect of the density is that the cosmological expansion rate depends on the total mass-energy density associated with a given temperature. For cosmological temperatures much above $10^4 K$, the energy density of radiation exceeds the mass-energy density of the baryon gas. Thus, during Big Bang Nucleosynthesis, we need the radiation density as well as the baryon density. The baryon density determines the density of the nuclei and thus their interaction rates, and the radiation density controls the expansion rate of the universe at those times. The density of radiation is just proportional to the number of types of radiation. Thus, the density of radiation is not a free parameter if we know how many types of relativistic particles exist when Big Bang Nucleosynthesis occurred.

Assuming that the allowed relativistic particles at $1 MeV$ are photons, e, μ, and τ neutrinos (and their antiparticles), and electrons (and positrons), Figure 1 shows the

BBN yields for a range in present ρ_b, going from less than that observed in galaxies to greater than that allowed by the observed large-scale dynamics of the universe. The ^4He yield is almost independent of the baryon density, with a very slight rise in the density due to the ability of nuclei to hold together at slightly higher temperatures and at higher densities, thus enabling nucleosynthesis to start slightly earlier, when the baryon to photon ratio is higher. No matter what assumptions one makes about the baryon density, it is clear that ^4He is predicted by Big Bang Nucleosynthesis to be around 1/4 of the mass of the universe.

As noted above, BBN yields all agree with observations using only one freely adjustable parameter, ρ_b. Recent attempts to circumvent this argument[28], by having variable n/p ratios coupled with density inhomogeneities inspired by a first order quark-hadron phase transition, fail in most cases to fit the Li and ^4He, even when numerous additional parameters are added and fine-tuned. In fact, it can be shown[29] that the observed abundance constraints yield such a robust solution that nucleosynthesis may constrain the quark-hadron phase transition more than the phase transition alters the cosmological conclusions.

This narrow range in baryon density for which agreement occurs is very interesting. Let us convert it into units of the critical cosmological density for the allowed range of Hubble expansion rates. From the Big Bang Nucleosynthesis constraints[8,9,10,11,27,28], the dimensionless baryon density, Ω_b, that fraction of the critical density that is in baryons, is less than 0.11 and greater than 0.02 for $0.4 \lesssim h_0 \lesssim 0.7$, where h_0 is the Hubble constant in units of $100 km/sec/Mpc$. The lower bound on h_o comes from direct observational limits and the upper bound from age of the universe constraints[30]. Note that the constraint on Ω_b means that the universe *cannot be closed with baryonic matter*. If the universe is truly at its critical density, then nonbaryonic matter is required. This argument has led to one of the major areas of research at the particle-cosmology interface, namely, the search for non-baryonic dark matter.

Another important conclusion regarding the allowed range in baryon density is that it is in very good agreement with the density implied from the dynamics of galaxies,

including their dark halos. An early version of this argument, using only deuterium, was described over ten years ago[31]. As time has gone on, the argument has strengthened, and the fact remains that galaxy dynamics and nucleosynthesis agree at about 6% of the critical density. Thus, if the universe is indeed at its critical density, as many of us believe, it requires most matter not to be associated with galaxies and their halos, as well as to be nonbaryonic. We will return to this point later.

Let us now look at the connection to N_ν. Remember that the yield of ^4He is very sensitive to the n/p ratio. The more types of relativistic particles, the greater the energy density at a given temperature, and thus, a faster cosmological expansion. A faster expansion yields the weak-interaction rates being exceeded by the cosmological expansion rate at an earlier, higher temperature; thus, the weak interaction drops out of equilibrium sooner, yielding a higher n/p ratio. It also yields less time between dropping out of equilibrium and nucleosynthesis at $10^9 K$, which gives less time for neutrons to change into protons, thus also increasing the n/p ratio. A higher n/p ratio yields more ^4He. Quark-hadron induced variations[28] in the standard model *also* yield higher ^4He for higher values of Ω_b. Thus, such variants still support the constraint on the number of relativistic species.[29]

In the standard calculation we allowed for photons, electrons, and the three known neutrino species (and their antiparticles). However, by doing the calculation (see Figure 2) for additional species of neutrinos, we can see when ^4He yields exceed observational limits while still yielding a density consistent with the ρ_b bounds from ^2H, ^3He, and now ^7Li. (The new ^7Li value gives approximately the same constraint on ρ_b as the others, thus strengthening the conclusion.) The bound on ^4He comes from observations of helium in many different objects in the universe. However, since ^4He is not only produced in the Big Bang but in stars as well, it is important to estimate what part of the helium in some astronomical object is primordial—from the Big Bang—and what part is due to stellar production after the Big Bang. The pioneering work of the Peimberts[32] showing that ^4He varies with oxygen has now been supplemented by examination of how ^4He varies with nitrogen[33] and carbon.[34] The observations have also been systematically
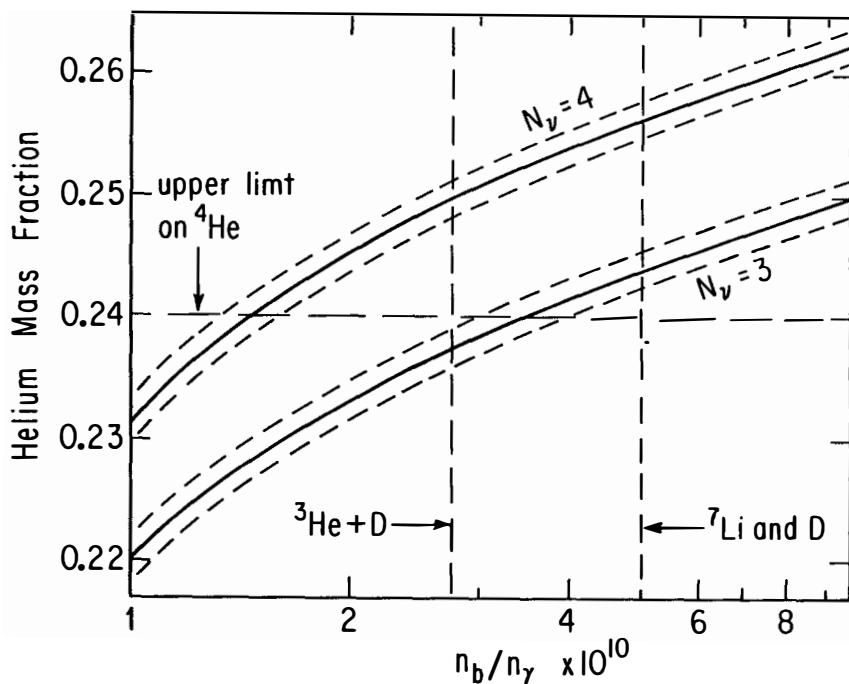
Figure 2. Helium mass fraction versus the cosmological baryon-to-photon ratio. The vertical line is the lower bound on this ratio from considerations of ^2H and ^3He (see Yang *et al.*) (Using ^7Li as a constraint would move the vertical line only slightly to the left.) The horizontal line is the current upper bound of 0.24. The width of the lines for $N_\nu = 3$ and 4 is due to $\tau_n = 890 \pm 4s$. Note that $N_\nu = 4$ appears to be excluded barring a systematic error upward in Y_p which would be contrary to current systematic trends.

reexamined by Pagel[33]. The conclusions of Pagel[33], Steigman *et al.*[34] and Walker *et al.*[10] all agree that the ^4He mass fraction, Y_p, extrapolated to zero heavy elements, whether using N, O, or C, is $Y_p \sim 0.23$ with an upper bound of 0.24. This has been reviewed in detail at this meeting by Keith Olive.[35]

The other major uncertainty in the ^4He production used to be the neutron lifetime. However, the new world average of $\tau_n = 890 \pm 4s(\tau_{1/2} = 10.3 \ min)$ is dominated by the dramatic results of Mampe *et al.*[36] using a neutron bottle. This new result is quite consistent with a new counting measurement of Byrne *et al.*[37] and within the errors of the previous world average of $896 \pm 10s$ and is also consistent with the precise C_A/C_V measurements from PERKEO[38 and others. Thus, the old ranges of $10.4 \pm 0.2 \ min$, used for the half-life in calculations,[39,8] seem to have converged towards the lower side. The convergence means that, instead of the previous broad bands for each neutrino flavour, we obtain relatively narrow bands (see Figure 2). Note that $N_\nu = 4$ is excluded. In fact, the upper limit is now $N_\nu < 3.4$.[9,10]

The recent verification of this cosmological standard model prediction by LEP, $N_\nu = 2.96 \pm 0.14$, from the ALEPH, DELPHI, L3 and OPAL collaborations presented elsewhere in this volume as well as the SLC results, thus, experimentally confirms our confidence in the Big Bang. (However, we should also remember that LEP and cosmology are sensitive to different things.[40] Cosmology counts all relativistic degrees of freedom for $m_x \lesssim 10 MeV$ with $m_x \lesssim 45 GeV$.

While ν_e and ν_μ are obviously counted equally in both situations, a curious loophole exists for ν_τ since the current experimental limit $m_\tau < 35 MeV$ could allow it not to contribute as a full neutrino in the cosmology argument[41]. It might also be noted that now that we know $N_\nu = 3$, we can turn the argument around and use LEP to predict the primordial helium abundance ($\sim 24\%$) or use limits on ^4He to give an additional upper limit on Ω_b (also $\lesssim 0.10$). Thus, LEP strengthens the argument that we need non-baryonic dark matter if $\Omega = 1$. In fact, note also that with $N_\nu = 3$, if Y_p is ever proven to be less than ~ 0.235, standard BBN is in difficulty. Similar difficulties occur if Li/H is ever found below $\sim 10^{-10}$. In other words, BBN is a falsifiable theory. (The

same cannot be said for many other astrophysical theories.)

Let us now put the nucleosynthetic argument on Ω_b into context.

DARK MATTER

The arguments requiring some sort of dark matter fall into two separate and quite distinct areas. First are the arguments using Newtonian mechanics applied to various astronomical systems that show that there is more matter present than the amount that is shining. These arguments are summarized in the first part of Table 1. It should be noted that these arguments reliably demonstrate that galactic halos seem to have a mass ~ 10 times the visible mass. At this Moriond meeting, the traditional dynamical arguments were supplemented by gravitational lens arguments by Petrossian[42] and Burke,[43] which also yielded relatively high mass-to-light ratios comparable to the dynamic masses implied on the same scales.

Note however that Big Bang Nucleosynthesis requires that the bulk of the baryons in the universe are dark since $\Omega_{vis} << \Omega_b$. Thus, the dark halos could in principle be baryonic[19]. Recently arguments (c.f. Bertschinger[44]) on very large scales[45] (bigger than cluster of galaxies) hint that Ω on those scales is indeed greater than Ω_b, thus forcing us to need non-baryonic matter. However, until the uncertainties in these arguments are fully investigated, we must look at the inflation paradigm for "proof" of $\Omega > \Omega_b$.

This paradigm is the argument that the only long-lived natural value for Ω is unity, and that inflation[46] or something like it provided the early universe with the mechanism to achieve that value and thereby solve the flatness and smoothness problems. (Inflation was much discussed at this meeting, so I will not bother to expand on it here. However, remember that up to now our need for exotica is dependent on inflation and Big Bang Nucleosythesis and not on the existence of dark galatic halos.)

Table 2 summarizes both the baryonic and non-baryonic dark matter candidates. Some baryonic dark matter must exist since we know that the lower bound from Big Bang Nucleosynthesis is greater than the upper limits on the amount of visible matter in the universe. However, we do not know what form this baryonic dark matter is in.

TABLE I

"OBSERVED" DENSITIES

$$\left[\Omega \equiv \rho/\rho_c \text{ where } \rho_c = 2 \cdot 10^{-29} h_0^2 \text{g/cm}^3 \text{and } h_o \equiv \frac{\text{Ho}}{100 \text{ km/sec/mpc}}\right]$$

Newtonian Mechanics
(cf. Faber and Gallagher[65])

Visible
$\Omega \sim 0.007$
(factor of 2 accuracy)

Binaries
Small groups
Extended flat relation curves
$\Omega \sim 0.07$
(factor of 2 accuracy)

Clusters
Gravitational lenses
$\Omega \sim 0.1 \text{to} 0.3$

Big Bang Nucleosynthesis (with $t_u \gtrsim 10^{10}$yrs.)
(c.f. Walker *et al.*[10] and ref. therein)
$\Omega_b = 0.065 \pm 0.045$

Preliminary Large Scale Studies

IRAS red shift study and peculiar velocities
(Ref. [45])
$\Omega \gtrsim 0.3$

Density redshift counts
(Loh and Spillar[66])
$\Omega \sim 1 \pm 0.6$

Inflation Paradigm
(Guth[46])
$\Omega = 1$

TABLE II

"DARK MATTER CANDIDATES"

Baryonic (BDM)

Brown Dwarfs and/or Jupiters	$M \lesssim 0.08 M_\odot$
Blackholes	$M \gtrsim 1 M_\odot$
Hot intergalactic gas	$M \sim 1 GeV, \; (T \sim 10^6 \, K)$
Failed galaxies	$M \gtrsim 10^5 M_\odot$

Non Baryonic

Hot (HDM)

Low Mass Neutrinos	$m_\nu \sim 20 \pm 10 eV$

Cold (CDM)

Massive Neutrinos	$m_\nu \sim 3 GeV \; (\gtrsim 45 GeV)^*$
WIMPS, Lightest Supersymmetric Particle (Photino, Gravitino, Sneutrino)	$m_{susy} \sim 4 GeV \; (\gtrsim 15 GeV)^*$
Axions	$m_a \sim 10^{-5} eV$
Planetary mass black holes	$M \sim 10^{15} g - 10^{30} g$
Quark nuggets	$M \sim 10^{15} g$
Topological debris (monopoles higher dimensional knots, balls of wall, etc.)	$M \gtrsim 10^{16} GeV$

* After LEP, etc.

It could be either in condensed objects in the halo, such as brown dwarfs and jupiters (objects with $\lesssim 0.08 M_\odot$ so they are not bright shining stars), or in black holes (which at the time of nucleosynthesis would have been baryons); c.f. Hegyi and Olive[47]. (Recently, Mathews and Schramm[48] have presented a galactic evolution model which puts large amounts of early objects into the halo.) Or, if the baryonic dark matter is not in the halo, it could be in hot intergalactic gas, hot enough not to show absorption lines in the Gunn-Peterson test, but not so hot as to be seen in the x-rays. Evidence for some hot gas is found in clusters of galaxies. However, the amount of gas in clusters would not be enough to make up the entire missing baryonic matter. Another possible hiding place for the dark baryons would be failed galaxies, large clumps of baryons that condense gravitationally but did not produce stars. Such clumps are predicted in galaxy formation scenarios that include large amounts of biasing where only some fraction of the clumps shine.

Non-baryonic matter can be divided following Bond and Szalay[49] into two major categories for cosmological purposes: hot dark matter (HDM) and cold dark matter (CDM). Hot dark matter is matter that is relativistic until just before the epoch of galaxy formation, the best example being low mass neutrinos with $m_\nu \sim 20 eV$. (Remember, $\Omega_\nu \sim \frac{m_\nu(ev)}{100 h_0^2}$).

Cold dark matter is matter that is moving slowly at the epoch of galaxy formation. Because it is moving slowly, it can clump on very small scales, whereas HDM tends to have more difficulty in being confined on small scales. Examples of CDM could be massive neutrino-like particles with masses greater than several GeV or the lightest super-symmetric particle which is presumed to be stable and might also have masses of several GeV (see detailed discussion of ref. [50]. Following Michael Turner, all such weakly interacting massive particles are called "WIMPS." Axions, while very light, would also be moving very slowly[51] and, thus, would clump on small scales. Or, one could also go to non-elementary particle candidates, such as planetary mass blackholes[52] or quark nuggets of strange quark matter, also found at the quark-hadron transition. Another possibility would be any sort of massive toplogical remnant left

over from some early phase transition. Note that CDM would clump in halos, thus requiring the dark baryonic matter to be out between galaxies, whereas HDM would allow baryonic halos.

When thinking about dark matter candidates, one should remember the basic work of Zeldovich,[53], later duplicated by Lee and Weinberg[54] and others,[55] which showed for a weakly interacting particle that one can obtain closure densities, either if the particle is very light, $\sim 20eV$, or if the particle is very massive, $\sim 3GeV$. This occurs because, if the particle is much lighter than the decoupling temperature, then its number density is the number density of photons (to within spin factors and small corrections), and so the mass density is in direct proportion to the particle mass, since the number density is fixed. However, if the mass of the particle is much greater than the decoupling temperature, then annihilations will deplete the particle number. Thus, as the temperature of the expanding universe drops below the rest mass of the particle, the number is depleted via annihilations. For normal weakly interacting particles, decoupling occurs at a temperature of $\sim 1MeV$, so higher mass particles are depleted. It should also be noted that the curve of density versus particle mass turns over again (see Figure 3) once the mass of the WIMP exceeds the mass of the coupling boson[56,57,58] so that the annihilation cross section varies as $\frac{1}{E^2}$, independent of the boson mass. In this latter case, $\Omega = 1$ can be obtained for $M_x \sim 1TeV \sim (3K \times M_{Planck})^{1/2}$, where $3K$ and M_{Planck} are the only energy scales left in the calculation (see Figure 3).

A few years ago the preferred candidate particle was probably a few GeV mass WIMP. However, LEP's lack of discovery of any new particle coupling to the Z° with $M_x \lesssim 45GeV$ clearly eliminates that candidate[59,60] (see Figures 4A and 4B). In fact, LEP also tells us that any particle in this mass range must have a coupling $\lesssim 10\%$ of the coupling of ν's to the Z°, or it would have shown up in the N_ν experiments. The consequences of this for $\Omega = 1$ dark matter are shown in Figures 4A and 4B for both Dirac (s-wave) and Majorana (p-wave) particles. Dirac particles are further constrained by the lack of detection in the ^{76}Ge experiments of Caldwell *et al.*[61]. The possibility of some other WIMP not coupling to the Z° is constrained by the non-detection of other
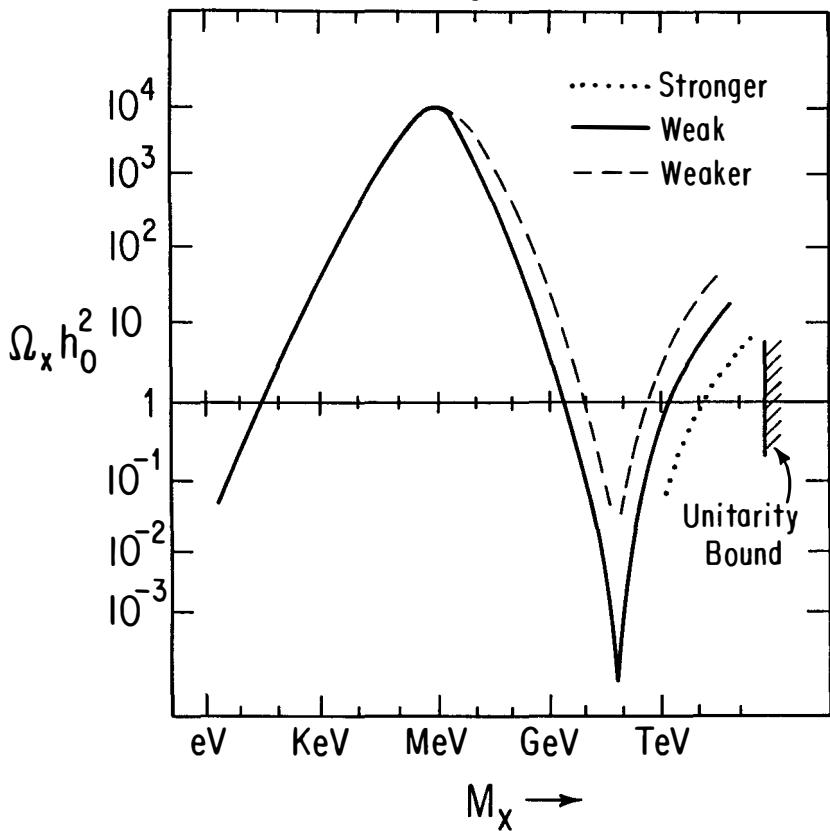
Figure 3. $\Omega_x h_0^2$ versus M_x for weakly interacting particles showing three crossings of $\Omega h_0^2 = 1$. Note also how curve shifts at high M_x for intereactions weaker or stronger than normal weak interaction (where normal weak is that of neutrino coupling through $\ddot{Z}^0$). Extreme strong couplings reach a unitarity limit at $M_x \sim 340 TeV$.
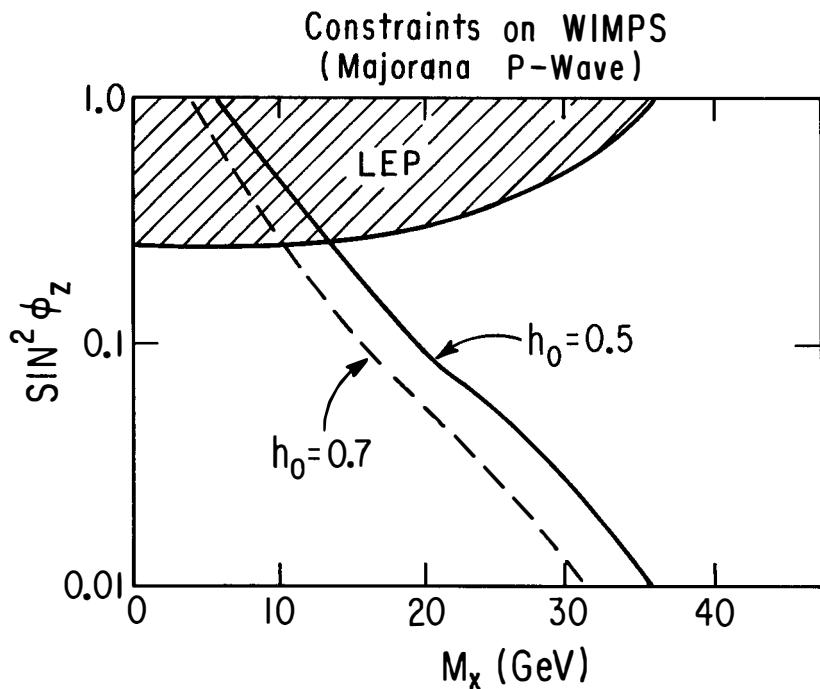
Figure 4 A. Constraints on WIMPS of mass M_x versus $sin^2\phi_z$, the relative coupling to the Z^0. The constraints are shown assuming Majorana particles (p-wave interactions). The diagonal lines show the combinations of M_x and $sin^2\phi_z$ that yield $\Omega = 1$. The cross-hatched region is what is ruled out by the current LEP results. Note that $\Omega = 1$ with $h_0 = 0.5$ is possible only if $M_x \gtrsim 15 GeV$ and $sin^2\phi_z < 0.3$. The new LEP run should lower this bound on $sin^2\phi$ to $\lesssim 0.1$.

Constraints on WIMPS
(Dirac S-Wave)
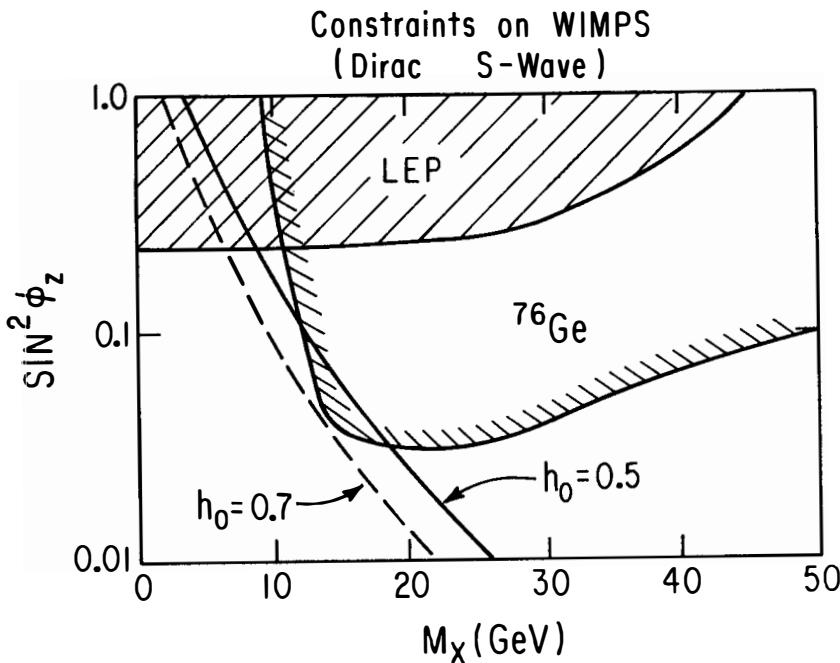
Figure 4B. This is the same as 4A but for Dirac particles (s-wave interactions), the ^{76}Ge region is that ruled out by the Caldwell *et al.* double-β decay style experiments. Note that while a small window for $\Omega = 1$, $h_0 = 0.5$, currently exists for $M_x \sim 10 GeV$, the combination of future ^{76}Ge experiments plus the new LEP run should eliminate this and leave only $M_x \gtrsim 20 GeV$ and $sin^2 \phi \lesssim 0.03$.

bosons, including squarks, sleptons and/or a Z', at $UA-2$ and CDF, as reported in the particle sessions of this meeting. Thus, with the exception of a few minor loopholes, whether the particle is supersymmetric or not, it is required to have an interaction weaker than weak and/or have a mass greater than about $20GeV$. This is discussed in detail in Ellis *et al.*[60] and also by Krauss[62] and Olive[35] in their talks at this meeting. Future dark matter searches should thus focus on more massive and more weakly interacting particles.

Also, as Dimopoulos[56] has emphasized, the next appealing crossing of $\Omega = 1$ (see Figure 3) is $\gtrsim 1TeV$ (but, in any case, $\lesssim 340TeV$ from the unitarity bound of Griest, Kamionkowski and Turner[58]; see also Hut and Olive[63] for limit with matter-antimatter assymmetry), which can be probed by SSC and LHC as well as by underground detectors. Thus, after LEP, the favoured CDM particle candidate is either a $10^{-5}eV$ axion or a gaugino with a mass of many tens of GeV, or a TeV mass particle. Of course, an HDM ν_τ with $m_{\nu_\tau} \sim 20 \pm 10eV$ is still a fine candidate as long as galaxy formation proceeds by some mechanism other than adiabatic gaussian matter fluctuations[64]

CONCLUSION

Particle experiments have tested the standard cosmological model, the Big Bang, in almost as dramatic a fashion as they have tested the standard particle model, $SU_3 \times SU_2 \times U_1$. The result is a continued confidence in the Big Bang and in the standard model conclusion that $\Omega_b \sim 0.06$. Particle experiments have even eliminated the favoured mass particles of a few years ago and force us to think in terms of heavier, more weakly interacting particles or revert to axions or a $20eV$ tau neutrino.

336

ACKNOWLEDGMENTS

References

1. J. Mather *et al.*, COBE preprint, Goddard *Astrophys. J.*, in press (1990).

2. R.A. Alpher, H. Bethe and G. Gamow *Phys. Rev.* **73**, 803 (1948).

3. R.A. Alpher, J.W. Follin and R.C. Herman *Phys. Rev.* **92**, 1347 (1953).

4. R. Taylor and F. Hoyle *Nature* **203**, 1108 (1964).

5. P.J.E. Peebles *Phys. Rev. Lett.* **16**, 410 (1966).

6. P. Wagoner, W.A. Fowler and F. Hoyle *Astrophys. J.* **148**, 3 (1967).

7. D.N. Schramm and R.V. Wagoner *Ann. Rev. of Nuc. Sci.* **27**, 37 (1977).
   K. Olive, D.N. Schramm, G. Steigman, M. Turner and J. Yang *Astrophys. J.* **246**, 557 (1981);
   A. Boesgaard and G. Steigman *Ann. Rev. of Astron. and Astrophys.* **23**, 319 (1985).

8. J. Yang, M. Turner, G. Steigman, D.N. Schramm and K. Olive *Astrophys. J.* **281**, 493 (1984) and references therein.

9. K. Olive, D.N. Schramm, G. Steigman and T. Walker *Phys. Lett B.* **236**, 454 (1990).

10. T. Walker, G. Steigman, D.N. Schramm and K. Olive *Astrophys. J.*, submitted (1990).

11. L. Kawano, D.N. Schramm and G. Steigman *Astrophys. J.* **327**, 750 (1988).

12. L. Kawano *Let's Go Early Universe*, FNAL (1988).

13. W. Fowler, J. Greenstein and F. Hoyle *Geophys. J.R.A.S.* **6**, 6 (1962).

14. C. Ryter, H. Reeves, E. Gradstajn and J. Audouze *Astron. and Astrophys.* **8**, 389 (1970).

15. J. Geiss and H. Reeves *Astron. and Astrophys.* **18**, 126 (1971);
   D. Black *Nature* **234**, 148 (1971).

16. J. Rogerson and D. York *Astrophys. J.* **186**, L95 (1973).

17. H. Reeves, J. Audouze, W.A. Fowler and D.N. Schramm *Astrophys. J.* **179**, 909 (1973).

18. R. Epstein, J. Lattimer and D.N. Schramm *Nature* **263**, 198 (1976).

338

19. J. R. Gott, III, J. Gunn, D.N. Schramm and B.M. Tinsley *Astrophys. J.* **194**, 543 (1974).

20. R.T. Rood, G. Steigman and B.M. Tinsley *Astrophys. J.* **207**, L57 (1976).

21. J. Yang, D.N. Schramm, G. Steigman and R.T. Rood *Astrophys. J.* **227**, 697 (1979).

22. T. Wilson, R.T. Rood and T. Bania *Proc. of the ESO Workshop on Primordial Healing*, ed. P. Shaver and D. Knuth (Garching: European Southern Observatory, 1983).

23. J. Spite and F. Spite *Astron. and Astrophys.* **115**, 357 (1982);

    R. Rebolo, P. Molaro and J. Beckman *Astron. and Astrophys.* **192**, 192 (1988);

    L. Hobbs and C. Pilachowski *Astrophys. J.* **326**, L23 (1988).

24. G. Steigman, D.N. Schramm and J. Gunn *Phys. Lett.* **66B**, 202 (1977).

25. V.F. Schvartzman *JETP Letters* **9**, 184 (1969);

    P.J.E. Peebles *Physical Cosmology* (Princeton University Press, 1971).

26. D.N. Schramm and L. Kawano *Nuc. Inst. and Methods* **A 284**, 84 (1989).

27. L. Krauss *et al.*, Yale University preprint (1989).

28. R. Scherrer, J. Applegate and C. Hogan *Phys. Rev. D 35*, 1151 (1987);

    C. Alcock, G. Fuller and G. Mathews *Astrophys. J.* **320**, 439 (1987);

    W.A. Fowler and R. Malaney *Astrophys. J.* **333** (1988) 14.

29. H. Kurki-Suonio, R. Matzner, K. Olive and D.N. Schramm *Astrophys. J.* **353**, 406 (1990);

    H. Reeves, K. Sato and M. Tarasawa, U. of Tokyo preprint (1989).

30. K. Freese and D.N. Schramm *Nucl. Phys.* **B233**, 167 (1984).

31. J. Gott, J. Gunn, D.N. Schramm and B.M. Tinsley *Astrophys. J.* **194**, 543 (1974).

32. M. Peimbert and S. Torres-Peimbert *Astrophys. J.* **193**, 327 (1974).

33. B. Pagel *ESO/CERN Proc.* (1990).

34. G. Steigman, D.M. Schramm and J. Gallagher *Comments on Astrophys.* **14**,

97 (1989).

35. K. Olive *Proc. Rencontres de Moriond* (1990).

36. W. Mampe, P. Ageron, C. Bates, J.M. Pendlebury and A. Steyerl *Phys. Rev. Lett.* **63**, 593 (1989).

37. J. Byrne *et al. Phys. Lett. B*, submitted (1990).

38. H. Abele, M. Arnould, H.A. Borel, J. Dohner, D. Dubbers, S. Freedman, J. Last and I. Reichert *Proc. Grenoble Workshop on Slow Neutrons* (1989).

39. G. Steigman, K. Olive, D.N. Schramm and M. Turner *Phys. Lett. B* **176**, 33 (1986).

40. D.N. Schramm and G. Steigman *Phys. Lett. B* **141**, 337 (1984).

41. E. Kolb and R. Scherrer *Phys. Rev. D* **25**, 1481 (1982);
    A. Chackrubat, E. Kolb and D.N. Schramm, in progress (1990).

42. V. Petrossian *Proc. Rencontres de Moriond* (1990).

43. B. Burke *Proc. Rencontres de Moriond* (1990).

44. E. Bertschinger *Proc. Rencontres de Moriond* (1990).

45. M. Strauss, M. Davis and A. Yahil, U.C. Berkeley preprint (1989);
    N. Kaiser and A. Stebbins, CITA preprint (1990);
    E. Bertschinger *et al.*, MIT preprint (1990);
    M. Rowan-Robinson and A. Yahil *Proc. Rencontres de Moriond* (1989).

46. A. Guth *Phys. Rev. D* **23**, 347 (1981).

47. D. Hegyi and K. Olive *Astrophys. J.* **303**, 56 (1986).

48. G. Mathews and D.N. Schramm *Astrophys. J.*, submitted (1990).

49. R. Bond and A. Szalay *Proc. Texas Relativistic Astrophysical Symposium, Austin, Texas* (1982).

50. M. Srednicki, R. Watkins and K. Olive *Nucl. Phys. B* **310**, 693 (1988).

51. M. Turner, F. Wilczek and A. Zee *Phys. Lett. B* **125** (1983) 35; **125**, 519 (1983).

52. D. Crawford and D.N. Schramm *Nature* **298**, 538 (1982).

53. Ya. Zeldovich *Adv. Astron. and Astrophys.* **3**, 241 (1965);

H.Y. Chin *Phys. Rev. Lett.* **17**, 712 (1966).

54. B. Lee and S. Weinberg *Phys. Rev. Lett.* **39**, 165 (1977).

55. C.P. Hut *Phys. Lett. B* **69**, 85 (1977);

    K. Sato and H. Koyayashi *Prog. Theor. Phys.* **58**, 1775 (1977).

56. S. Dimopoulos, R. Esmailzadeh, L. Hall and N. Tetradis *Nucl. Phys. B*, submitted (1990).

57. D. Brahm and L. Hall *Phys. Rev. D* **41**, 1067 (1990).

58. K. Griest, M. Kamionkowski and M. Turner, FNAL preprint (1990).

59. L. Krauss *Phys. Rev. Lett.*, in press (1990);

    K. Griest and J. Silk, U.C. Berkeley preprint (1990).

60. J. Ellis, D. Nanopoulos, L. Roskowski and D.N. Schramm *Phys. Lett. B*, in press (1990).

61. Caldwell *et al.*, *Proc. La Thuile*, ed. M. Greco (1990).

62. L. Krauss *Proc. Rencontres de Moriond* (1990).

63. P. Hut and K. Olive *Phys. Lett.* **87B**, 144 (1979).

64. D. N. Schramm, in *Particle Astrophysics and Experimental Issues*, ed. E. Norman (Singapore: World Scientific, 1989).

65. S.M. Faber and J.S. Gallagher *Ann. Rev. Astron. and Astrophys.* **17**, 135 (1979).

66. E. Loh and E. Spillar *Astrophys. J.* **329**, 24 (1988).

THE IMPLICATIONS OF Z WIDTH MEASUREMENTS FOR THE SEARCH FOR DARK MATTER

OR

THERE IS NO SUBSTITUTE FOR EXPERIMENTAL DATA

LAWRENCE M. KRAUSS[1]

CENTER FOR THEORETICAL PHYSICS AND DEPT. OF ASTRONOMY
SLOANE LABORATORY, YALE UNIVERSITY
NEW HAVEN, CT 06511

Abstract

I review the implications of the Z decay measurements at SLC and LEP for the search for dark matter: (a) The Z width measurements, when combined with non-accelerator data rule out Dirac neutrinos, Majorana neutrinos, and sneutrinos as WIMP dark matter, strongly constrain neutralinos, and rule out most explicit "cosmions"; (b) The limit on the number of neutrinos, when combined with BBN estimates, could provide evidence both for baryonic and non-baryonic dark matter; (c) direct searches for WIMPs will require new detectors, sensitive to spin dependent interactions on possibly heavy targets, with event rates which are 3-4 orders of magnitude smaller than those expected at present detectors.

INTRODUCTION: THE Z WIDTH REVIEWED:

In less than a year of running, measurements of Z decay parameters at LEP and also SLC have already significantly constrained our picture of the world. It is remarkable how even a small amount of unambiguous data about so fundamental a system can so strongly direct theory. In particular, the new limits on Z width constrain not only the number of light neutrino species in nature, but also the existence of new particles of mass up to 45 GeV which might be produced in Z decay. Among these are included many of the leading Weakly Interacting Massive Particle (WIMP) candidates for dark matter [1]. Those which remain viable will be far more challenging for experimentors to detect directly.

Before going into the detailed constraints, and because it is important for everything I will have to say here, I will first review in detail how one infers a limit on the number of neutrinos from the Z decay measurements made at SLC and LEP. Such an analysis is essential if one is to attempt to compare and combine the results from the different experiments described in this volume, and determine what are the limiting factors which may affect further measurements.

When Z bosons are produced at e+e- colliders, the shape of the Z resonance is most easily determined by observing the hadronic decays of the Z, which have by far the largest branching ratio, and so lead to the best statistics. By counting the number of decays as a function of energy, and carefully measuring the luminosity of the machine at each energy, the cross section for Z production can be mapped out, and compared to the standard Breit-Wigner form:

$$\sigma_h = \sigma_h^0 \frac{s\,\Gamma_z^2}{\left(s - M_z^2\right)^2 + \left(s/M_z^2\right)\Gamma_z^2}\left[1 + \delta_{rad}(s)\right] \tag{1}$$

Here, $\sqrt{s}$ is the center of mass energy, $[1+\delta(s)]$ accounts for initial state radiation, and the Breit-Wigner resonance form is then described by 3 parameters: M_z, Γ_z, and σ_h^0, the mass, the total width, and the height at the peak respectively. These are displayed in the figure below.
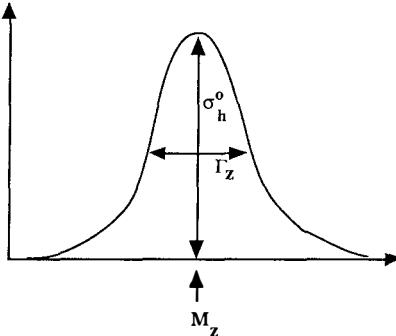
Not all these parameters are independent, however. In particular, in order that the integral under the curve give the total Z production cross section, the peak height σ^0 and the width Γ are related, so that:

$$\sigma_h^o = \frac{12\pi \Gamma_e \Gamma_h}{M_z^2 \Gamma_z^2} \tag{2}$$

where Γ_e and Γ_h are the Z partial decay widths into electrons and hadrons respectively.

The total Z width is **defined** to be: $\Gamma_z \equiv N_v \Gamma_v + 3\Gamma_e + \Gamma_h$ if lepton universality is assumed. Otherwise, the sum of partial widths into charged lepton pairs may be used in the definition. In any case, this defines what is meant by the "number" of neutrinos, i.e.

$$N_v \equiv \frac{1}{\Gamma_v}\left(\Gamma_z - 3\Gamma_e - \Gamma_h\right) \tag{3}$$

One's naive expectation might be that to extract a limit on the number of neutrinos one would first fit the measurements to the Breit-Wigner form (1), determine a best fit value for Γ_z and then use (3). However, since each new neutrino species contributes about 6% to the width, an uncertainty in the width determination of 6% corresponds to an uncertainty of one extra neutrino type. Instead, since σ_h^o depends *quadratically* on Γ_z (from (2)), an extra neutrino species contributes about 12% to the peak cross section value. Thus, assuming the uncertainty in the peak cross section determination is no worse than that in the width, one can get *twice* as good a limit on the number of neutrinos by using the fit for σ_h^o and then using (2) to limit the number of neutrinos than one can get by using Γ_z alone. Moreover, σ_h^o is defined in (2) in terms of *ratios* of widths, and is thus less sensitive to uncertainties in the widths from the standard model, such as from the top quark mass, or the uncertainty in $\sin^2\theta_w$. Thus, using (2), we can write

$$N_v \equiv \frac{\Gamma_e}{\Gamma_v}\left(\sqrt{\frac{12\pi}{M_z^2 \sigma_h^o}}\sqrt{\frac{\Gamma_h}{\Gamma_e}} - 3 - \frac{\Gamma_h}{\Gamma_e}\right), \tag{4}$$

which expresses N_v in terms of σ_h^o and in terms of the ratio Γ_h/Γ_e, which can either be extracted from the standard model, or more important, can be directly measured. In the latter case, the ratio Γ_e/Γ_v must be taken from theory. (There is no such thing as completely model independent bound on N_v.) The first limits on N_v essentially used (2), (or (4))[2]. More recently the LEP groups have been quoting limits from (3), and combined limits from both (3) and (4) [3].

It is important, when comparing the results of different experiments, to make sure that the limits which are quoted are derived the same way, and also to attempt to use the firmest bounds. As an example, compare below different preliminary limits quoted at Moriond by two of the LEP experiments (Because these were preliminary they may not appear in the articles from the experimentors in this proceedings [3]):

Experiment	Γ_z	σ_h^o	Combined Limit
ALEPH	3.44±0.37	not quoted	3.03±0.15
L3	3.32±0.32	3.28±0.18	3.29±0.17

As expected, the errors on the limits obtained using Γ_z are about twice as big as those obtained using σ_h. Also note how the Γ_z bound can be more susceptible to systematic errors in partial rates. Finally, note that the combined limit is at present completely dominated by the σ_h constraint.

Before quoting the current fits and combined limit on N_v, I want to mention one aspect of the above which will impact on limits I will later discuss. As mentioned, sensitivity of σ_h^0 to N_v comes from the quadratic dependence on Γ_z. This implies however that any bound dominated by the peak cross section constraint has an inherent insensitivity to unstable particles. For example, Z decay into a heavy unstable fourth generation neutrino with couplings and decays to the third generation could effectively contribute to the measured hadronic partial width, and thus contribute to both the numerator and denominator of (2). The sensitivity to such an extra neutrino could then be reduced by a factor of 2 compared to the quoted limits, which strictly apply to stable particles, unless they are supplemented by other information from the experiments.

Caveats aside, finally, what is the current combined limit on N_v from all the experiments. Below, I display the compilation I performed based on the initial LEP preprints[2], and also the most recent results, obtained from presentations at Moriond[3], and later at Wisconsin[4], being as careful as possible to quote the limits derived from the same analysis in each experiment.

Experiment	N_v Limit (Oct 89)	N_v Limit (Mar 90)
Mark II	2.8 ± 0.6	same
ALEPH	3.27 ± 0.30	3.03 ± 0.15
DELPHI	$2.4 \pm 0.4 \pm 0.5$	3.05 ± 0.28
L3	3.42 ± 0.48	3.11 ± 0.17[4]
OPAL	3.12 ± 0.42	3.09 ± 0.19
Average	**3.13 ± 0.20**	**3.06 ± 0.09**

Finally, after having gone to all the trouble to explain these results, what is the point, i.e. what new information about dark matter may we glean? First, as I have stressed, N_v is *defined* by (3) or (4). As such, it can represent more than just new light or massless neutrino species. In particular, any new undetected particle which couples pairwise to the Z and is lighter than half its mass can contribute to N_v. This includes many of the current WIMP candidates.

NEW CONSTRAINTS ON WIMPS, ETC:

We can determine the contribution of any new particle to the Z width as follows. Fermions (mass M and coupling $L = [g/2 \cos \theta_w](G_v \bar{\psi}\gamma^\mu \psi + G_a \bar{\psi}\gamma^\mu \gamma^5 \psi)$ to the Z) contribute to the Z width as follows:

$$\Gamma = \frac{G_F M_Z^3}{6\sqrt{2}\pi}\left[\sqrt{1-4M^2/M_Z^2}\right]\left\{(G_v^2 + G_a^2)[1-M^2/M_Z^2] + 3(G_v^2 - G_a^2)M^2/M_Z^2\right\} \quad (5)$$

Complex scalars with mass M and $L = [ig K/2 \cos \theta_w](\varphi^* \partial^\mu \varphi - (\partial^\mu \varphi^*)\varphi)$ yield

$$\Gamma = \frac{G_F K^2 M_Z^3}{12\sqrt{2}\pi}\left[1-4M^2/M_Z^2\right]^{3/2} \quad (6)$$

One can determine what fraction of a neutrino these might mimic, by comparing these widths to that for a massless v.

$$\Gamma(Z^0 \to \bar{v}v) = G_F M_Z^3/12\sqrt{2}\pi \approx 165 \, MeV. \quad (7)$$

Then, by comparing the predictions to the 2σ upper bound on N_v of 3.53 (Oct.) or 3.25 (March), one can derive limits on new particle masses and couplings of new particles so that this bound is not exceeded. *For particles which couple with the same strength as neutrinos*, one finds[1]:

Particle Type	$N_v < 3.53$ (2σ)	$N_v < 3.25$ (2σ)	
fermion	$m \geq O(30)$ GeV	$m \geq O(40)$ GeV	(8)
scalar	$m \geq 0$ GeV	$m \geq O(30)$ GeV	

We shall see that most of the new constraints are relatively insensitive to the precise bound on N_v which is used.

(a) **Dirac Neutrinos:** The prototypical WIMPs, heavy Dirac neutrinos were the first elementary particles whose non-thermal remnant abundance was calculated in a manner identical to that used previously to calculate successfully the remnant abundance of light elements such as Helium produced in the big bang. The idea is simple. Heavy neutrinos maintain an equilibrium density as long as their annihilation rate exceeds the expansion rate (assuming no particle antiparticle asymmetry). After this time, the ratio of neutrinos to thermal photons in the universe is frozen in at the Boltzmann factor appropriate to the temperature at freeze-out:

$$\frac{n_v}{n_\gamma} \approx exp\left[-\frac{m_v}{T_{freeze\ out.}}\right] \tag{9}$$

$T_{freeze\ out}$ is fixed by the annihilation cross section. Since $<\sigma v> \approx G_F^2 m^2$ for Dirac neutrinos, it was shown in 1977[5] that if $m_v \approx 2$ GeV, (9) would imply a closure density of neutrinos in the universe today. Moreover, since the cross section increases quadratically with mass, *heavier* neutrinos will annihilate more efficiently, and so will have a *smaller* remnant mass density today. It is this dynamical coupling between weak interaction strength and GeV mass scale that makes WIMPs natural candidates for dark matter. Of course, if there is a particle-antiparticle asymmetry, this coupling is removed, and heavy neutrinos of any mass greater than 2 GeV could have a closure density today.

The LEP limits require the mass of a new stable Dirac neutrino to be greater than about 40 GeV. When this is incorporated in the calculations described above one finds that the remnant abundance of such particles, in the absence of an asymmetry, is far too small to make up all of even the galactic halo dark matter for masses between 40 GeV and about 2 TeV. Thus, the LEP results alone, in the absence of an asymmetry, rule out WIMP scale Dirac neutrinos as dark matter.

Fortunately, however, the LEP results can be supplemented by non-accelerator experiments, which can directly probe for a flux of dark matter WIMPS at the earth's surface. Low background Ge detectors are sensitive to ionization caused by energy deposited in elastic scattering off of nuclei for masses in excess of about 12 GeV, and the nonexistence of a signal above background[6] rules out Dirac neutrinos as galactic halo dark matter for 12 GeV $\leq m \leq 2$ TeV. This bound, which would require $m \leq 12$ GeV, is exactly complementary to the LEP bound, which requires $m \geq O(40)$ GeV. **The combination completely rules out Dirac neutrinos as WIMPs.** (Note: neutrinos with mass in excess of 2 TeV, which would be super-weakly interacting with normal

matter, and thus are un-WIMPlike, remain viable, if they have a remnant asymmetry.)

(b) Majorana neutrinos: Majorana neutrinos are identical to their antiparticles and therefore cannot have a primordial asymmetry. Thus, their relic abundance is entirely determined by annihilations in the early universe. For Majorana neutrinos however, $<\sigma v> \approx G_F^2 \, \mathbf{p}^2$, and so annihilation is suppressed for non-relativistic particles, leading to a slightly higher value for the mass resulting in a closure density, of about 5 GeV[7]. The relic density falls above this roughly as $m^{-5/3}$. Thus, the newest LEP limits imply that the fraction of closure density (Ω) in Majorana neutrinos must be less than about .01, or about equal to the visible mass density in the universe today. **Thus, the result (8) by itself rules out Majorana neutrinos as WIMPs.**

(c) SUSY WIMPs: Low energy supersymmetry provides the most compelling WIMP dark matter candidates. If the SUSY breaking scale M is tied to the weak scale, then the lightest SUSY particle (LSP) can get a mass of order $\alpha M \approx O(GeV)$. Morever, since the other SUSY partners of ordinary matter which can mediate in scattering and annihiltion processes have masses of order $M \approx M_w$, the LSP can have an interaction strength which is comparable to that of an ordinary massive neutrino. This combination makes it a natural WIMP![8] Which particle is lightest is model-dependent. Those often discussed include the sneutrino---the partner of the neutrino, and a "neutralino" ---the fermion partner of a linear combination of the photon, the Z, and the two Higges present in SUSY models. I discuss these separately below:

(1) Scalar Neutrino: "Sneutrinos" couple to the Z with the same strength as neutrinos, and hence, from (2) will contribute $\approx 1/2$ as much to the Z width, for the same mass. Hence, while the initial LEP results did not rule out such a particle accessible in Z decay, the newest results imply $m \geq O(30)$ GeV. Again, non-accelerator data provides a strong complimentary upper bound on its mass. The WIMP direct detection experiments which limit neutrinos also limit sneutrinos to be lighter than O(12) GeV. This limit is supplemented by data from indirect detection using proton decay detectors. Sneutrinos are not only efficiently captured by the sun and earth[9], but they annihilate into light neutrinos, yielding a signal in proton decay detectors which is not seen, implying $m \leq O(3\text{-}5)$ GeV[10]. **Combining limits, sneutrinos are ruled out as WIMPs.**

(2) Neutralinos: In low energy SUSY there are four neutral majorana fermion "partners" of ordinary matter--the photino, the zino, and two Higgsinos--- which are expected to be among the lightest states[8]. The states will mix in general, so that the mass eigenstates will be linear combinations of the weak eigenstates. In the minimal model the masses and couplings of the inos depend on four parameters: M_1 and M_2, the gauge fermion mass terms, the higgsino mass parameter μ, and the ratio of the two Higgs doublet expectation values v_2/v_1. If the model is unified at some scale then M_1 and M_2 are related by $M_1 = 5/3 \, \tan^2\theta_W M_2$, leaving three free parameters. Because of this larger freedom in model building, the constraints derivable from the Z decay width are less easily stated. For example, the pure neutralino states tend to decouple from Z decay. A pure photino or a pure Zino have no diagonal Z couplings, since their boson partners have none. It also turns out that a light pure Higgsino tends to be the linear combination which has vanishing Z coupling.

Nevertheless, some non-trivial constraints on the parameter space are derivable. Even in the cases where the LSP decouples from Z decay, the other neutralinos can contribute to the Z width[11], remembering of course that the heavier ones can decay and therefore may contribute 1/2 as much to the inferred invisible width (see above). Moreover, "charginos", the fermionic partners of the charged gauge bosons and Higgses, can give even larger contributions to the Z width[12]. These particles can have triplet weak isospin quantum numbers. For a charged gaugino, the Z decay width can be larger than that for 4 neutrinos[12]. It is conventional to present constraints as curves in M-μ space, for a fixed value of v_1/v_2. This ratio reflects the origin of electroweak symmetry breaking. If the top quark mass term drives spontaneous symmetry breaking in the higgs sector, then v_1/v_2 must be greater than unity (in the limit $v_1/v_{2=1}$ the LSP tends to pure photino or Higgsino and thus decouples from the Z), given the current top quark mass lower limit of 89 GeV. As v_1/v_2 increases, the Z width constraints become more powerful.

To derive constraints on the Z width, one diagonalizes the neutralino and chargino mass matrices, finds the weak couplings of the mass eigenstates, and plugs them into (5)[11,12,13]. The largest decay branching ratios come from the charginos (see above), so that the lightest chargino is constrained to have a m$\geq$O(M$_Z$/2) before its contribution to the Z width is below the present upper limit. The range in M-μ space ruled out by the requirement that m > O(40) GeV[12], is shown schematically below for two values of β. Also shown is the region of this parameter space which is ruled out by the requirement that the total Z decay width into neutralinos be less than .6 $\Gamma_Z(\nu)$[11]. Finally, I display estimates[14] of the cosmological mass density in the neutralino LSP, with the solid line for $\Omega=1$ and the dashed curve for $\Omega=0.1$. Squark and slepton and top masses $\approx$ O(100) GeV were assumed. Presumably the mass density of light neutralino WIMPs should lie between these two values. As these figures show, a significant region of the M-μ plane is constrained by the Z width limit.
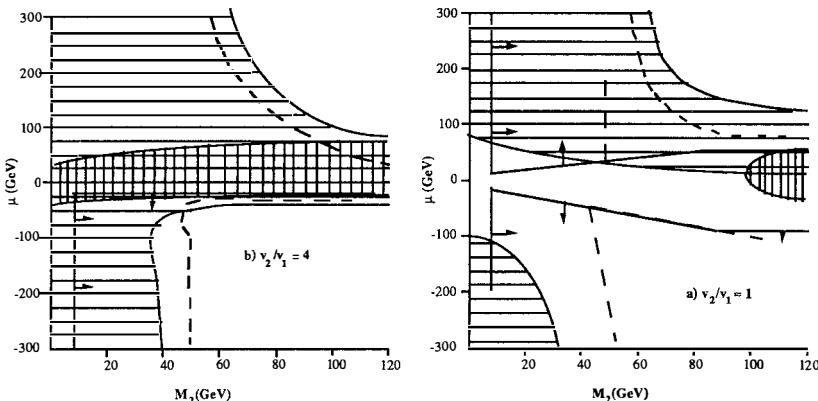
For $v_2/v_1 > 1$, much of the region M,$|\mu| <$ O(30) GeV is ruled out, and most of the remaining

cosmologically interesting range involves large positive M and μ values. In fact[15], a large parameter range exists for which a cosmologically relevant LSP is relatively heavy and is nearly pure bino or Higgsino. For $v_2/v_1 > 1$, the Z results are pushing neutralino WIMPs in this direction.

The Z data thus suggests that the LSP is probably heavier than O(20) GeV, and is not likely to be a light pure photino, but rather a mix of neutralinos, whose couplings with ordinary matter might include significant scalar contributions.

(d)**Exotica/cosmions**: Other than these WIMP candidates from particle physics, exotic objects have been proposed for purely astrophysical reasons. One of these is a "cosmion", a fixed abundance of which inside the sun might, for a mass of 4-10 GeV, lower the core temperature and reduce the solar neutrino flux.[16] This caused interest when it was suggested[17] that WIMPs might in principle be captured in the Sun over cosmological time to produce the required abundance, thus potentially solving both the dark matter and solar neutrino problems. However none of the standard WIMPs fit the severe requirements[18] which may make this solution appear contrived. Scattering cross sections 20-100 times larger than weak are required, as is the absence of annihilation inside the sun. However, in times of need ugliness is no obstacle, and theorists with time on their hands have produced exotic cosmion models. I describe here how the Z decay data can rule out many explicit models which previously appeared viable:

(a) Magnino, and Neutrino-Higgs: Raby and West proposed several cosmion models[19]. Each of these involved a Dirac neutrino with mass 5-10 GeV as a cosmion (with an additional interaction mediated either by a large magnetic moment, or by light Higgs exchange). The Z limit on extra massive Dirac neutrinos (and on a light Higgs) rules out both these possibilities.

(b) SUSY cosmions[20] : This model requires SUSY, and also a 1-2 Higgs. The only viable model proposed involved $v_2/v_1 \approx 1$, M=80-105, and μ=130-150 and a lightest chargino mass $\approx$30-40 GeV. These are ruled out by the Z decay data as discussed above, as is a light Higgs particle.

(c) E_6 cosmions[21]: This model proposed new v's in a 27 of E_6 with no Z couplings, but couplings to a new Z' particle. While such v's would not be produced in Z decay, there also exists one new doublet v per family. One might naively expect their mass to be about equal to that of the 5-10 GeV singlet cosmion. In this case this model would be ruled out by the Z width limit. Also, Z limits on light Z' particles are appearing which might also rule out such models.

(d) Colored Scalars: These models[22] are the least explicit, and hence difficult to rule out. But, all involve new colored scalars and heavy fermions. If they have standard weak quantum numbers, the Z data would then rule them out.

THE Z WIDTH AND BARYONIC DARK MATTER

It is by now very well known that the number of light neutrinos strongly impacts upon cosmology via primordial big bang nucleosynthesis. Traditionally, cosmologists have used the quantitative agreement between the predicted and the observationally inferred primordial abundances of light elements to limit the number of light neutrinos[23]. Now that this number has been well established experimentally via the Z width, one can hope to use it to further constrain other aspects of big bang cosmology. In particular, with the number of neutrinos known

(*assuming* no other light states which can affect the expansion rate during helium production, but which do not couple to the Z), the uncertainty in the predicted abundance of ^{4}He is the smallest of all the light elements. Below I display the BBN predictions for the fraction by mass of ^{4}He (Y_p), for 3 and 4 neutrinos as a function of the baryon to photon ratio η[24], including the 2σ uncertainty due to our uncertainty in the BBN reaction rates (dominated by the uncertainty in the neutron half life). Also shown for 3 neutrinos is the reduced 2σ uncertainty with the most recent neutron half life measurement included.
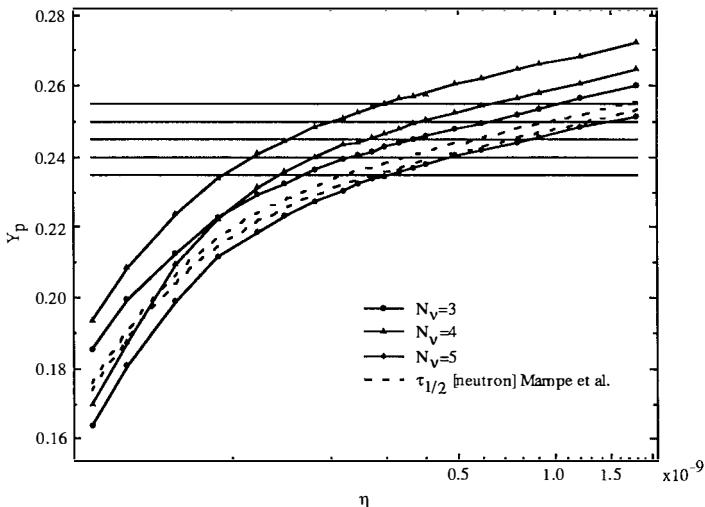
It is clear from the figure that if we can pin down the actual primordial abundance of ^{4}He we could place strong limits *from He alone* on the density of baryons in the universe today. The fraction of closure density in baryons (Ω_B) is determined from η as

$$\Omega_B = .0036 h_o^{-2} \ (T / 2.74 \ K)^3 \ (10^{10}\eta) \tag{10}$$

where T is the present microwave background temperature, and h_o is limited by measurements of the Hubble constant today to be between .4 and 1. For example, if we were to limit the primordial abundance to be between 23.5 and 24% for example, then, *assuming at most 3 light neutrino equivalents,* η would be constrained to be between 2.5-4.8 x 10^{-10}. This in turn would limit Ω_B to be between 0.01 and 0.1. The lower bound (obtained for the extreme value $h_o =1$, which leads to conflicts with limits on the age of the universe) is already suggestive that some dark matter must be baryonic. The upper limit is only marginally in agreement with the possibility that all dark matter inferred by virial estimates is baryonic. While such a narrow range for the primordial He abundance cannot be inferred from the present data[25], the Z width data, in combination with more careful analyses of big bang nucleosynthesis are bringing us closer definitively limiting the amount of baryonic dark matter in the universe.

CONCLUSIONS: IMPLICATIONS FOR DIRECT WIMP DETECTION

The Z width data effectively rules out all known WIMP candidates which interact coherently with total nuclear charge in low energy scattering processes. In addition, it suggests that the LSP, if it exists, is probably heavier than 10-20 GeV, and is not likely to be a pure photino. These have important implications for direct detection schemes, because they suggest that a viable WIMP is likely to be orthogonal to present detection schemes. Much of the thrust of ongoing WIMP detection experiments has been to develop sensitivity to probe for coherently interacting particles with mass less than 5-10 GeV. For Dirac neutrinos and sneutrinos, the LEP results now remove the need. Because they also rule out most cosmion models, that motivation for probing the light mass range is at least reduced. None of the present detectors is sensitive to the small rates which would result from the spin-dependent interactions, typical of neutralinos, and heavier neutralino and squark mass limits will make interaction rates even smaller. Trying to maximize the recoil energy will suggest going to heavier nuclear targets. Recent work[26] suggests that the rates in nuclear targets is suppressed compared to original estimates. The challenge to direct detection has increased! It may be that experimentors will have to wait first for a signal for the LSP at LEP or the Tevatron, in order to know how to optimize their detectors.

It is remarkable to have data which conclusively limits an area in which there has been so much speculation over the last decade. The first results from physics at the Z have provided powerful new limits, not just on the number of light families, but on the nature of dark matter. We can hope that the coming years may not only yield further constraints, but that a clear signal to guide us might emerge. If so, the identity of most of the mass of the universe could emerge from the data now being recorded at the Z pole.

REFERENCES

1. most of the limits I discuss here were derived in L.M. Krauss, Phys. Rev. Lett. **64**, 999 (1990)
2. L3 , Phys. Lett **B231**, 509 (1989); ALEPH, Phys. Lett **B231**, 519 (1989) ; OPAL, Phys. Lett **B231**, 530 (1989); DELPHI, Phys. Lett **B231**, 539 (1989); G.S Abrams et al. Phys. Rev. Lett. **63** 2173 (1989)
3. i.e. see L3, OPAL, ALEPH, and DELPHI reports, in this proceedings
4. All numbers quoted at Wisconsin were identical to those quoted at Moriond, with the exception of the L3 collaboration, which reported a change in their estimated luminosity, which reduced their quoted value on N_v by .18.
5. B.W. Lee and S. Weinberg, Phys. Rev. Lett. **39** 165 (1977); D. Dicus et al, Phys. Rev. Lett.**39**, 168 (1977)
6. D. Caldwell et al, Phys. Rev. Lett. **61**, 510 (1988); S. Ahlen et al, Phys. Lett. **B195**, 603 (1987) For a description of a more recent experiment with Si, which is now reporting limits in the range of 5 GeV, see B. Sadoulet et al, Astrophys. J. Lett. **324**, 75 (1988)
7. L. M. Krauss, Phys. Lett 128B, 37 (1983); E. Kolb, K. A. Olive, Phys. Rev **D33**, 1202

(1986)

8. for a review: J. Polchinski in *Inner Space, Outer Space,* ed. E. Kolb et al, University of Chicago Press (Chicago 1986) or see J. Ellis et al, Nucl. Phys. **B238**, 453 (1984)

9. L. M. Krauss, Harvard preprint HUTP-85/A008a (1985); J. Silk, K. Olive and M. Srednicki, Phys. Rev. Lett. **55**, 257 (1985); L. M. Krauss et al, Astrophys. J. **299**, 1001 (1985) : A. Gould, Astrophys. J.**321**, 560 (1987); ; L. M. Krauss, M. Srednicki and F. Wilczek, Phys. Rev. **D33**, 2079 (1986)

10. i.e. see K. Olive and M. Srednicki, Phys. Lett. **B205**, 553 (1988)

11. i.e. R. Barbieri et al, Phys. Lett. **195B,** 500 (1987); V. Barger et al, Phys. Rev. **D28**, 2912 (1983)

12. A. Bartl et al, Wien pub. HEPHY-PUB 522/89 and refs therein

13. I. Adachi et al. Phys. Lett. **B218**, 105 (1989)

14. i.e. see J. Ellis et al, in ref 12.

15. M. Srednicki and K. Olive, Minnesota pub. UMN-TH-801/89

16. J. Faulkner, R. Gilliland, Astrophys. J. **299**, 994 (1985) ; D. Spergel, W. Press, Astrophys. J. **294**, 663 (1985)

17. L. M. Krauss, Harvard preprint HUTP-85/A008a (1985); W. Press, D. Spergel, Astrophys. J. **296**, 673 (1985)

18. L. M. Krauss et al, Astrophys. J. **299**, 1001 (1985)

19. S. Raby and G. B.West, Nucl. Phys. **B292**, 793 (1987); Phys. Lett. **194B**, 557 (1987); Phys. Lett. **202B**, 47 (1988)

20. G.F. Guidice, E. Roulet, Phys. Lett. **219**, 309 (1988)

21. G.G. Ross, G. C. Segre, Phys. Lett. **197**, 45 (1987)

22. G. Gelmini, L. Hall, M.J. Lin, Nucl. Phys. **B281**, 726 (1987)

23. i.e. G. Steigman, D.N. Schramm, and J. E. Gunn, Phys. Lett. **B66**, 202 (1977)

24. This curve comes from L. M. Krauss, P. Romanelli, Yale Preprint #YCTP-P1-89, Astrophys. J. to appear July 20, 1990 issue.

25. See K. Olive, these proceedings.

26. i.e. A.F. Pacheco D. Strottman, Phys. Rev. **D40**, 2131 (1989); i.e. see J. Engel, P. Vogel, Phys. Rev. **D40**, 3132 (1989); and most recently, F. Iachello, L.M. Krauss, and G. Maino, Yale preprint YCTP-P14-89 (April 1990)

MASS AND LIGHT IN COSMIC STRUCTURES

FROM MASS TO LIGHT:
MODELS OF FORMATION AND EVOLUTION OF GALAXIES

B. Guiderdoni[1] and B. Rocca-Volmerange[1,2]
1) Institut d'Astrophysique de Paris, 98bis Boulevard Arago, F-75014 Paris
2) Université Paris-Sud, Bât. 333, F-91405 Orsay-Campus

ABSTRACT. The models of galaxy formation predict the *mass* distribution of galaxies from the linear fluctuation spectrum. So the predictions are not directly comparable with the observations which give the *luminosity* distribution, unless some ad hoc mass–to–light ratio is arbitrarily fixed. This procedure is far too crude and realistic models predicting consistent mass–to–light ratios are needed. We hereafter describe such a model of spectrophotometric evolution. Simple scenarios of evolution correctly reproduces the properties of nearby galaxies. As for distant galaxies, faint counts give a deep insight on the properties of the global population at high redshift. Nevertheless the straightforward interpretation of the data with the simple scenarios gives $\Omega_0 = 2q_0 \simeq 0.10$ which is inconsistent with the inflationary paradigm. In order to save $\Omega_0 = 2q_0 = 1$, one must introduce a new assumption, for instance strong number evolution or an additional population of low–surface–brightness galaxies. In this latter context, we discuss preliminary results of a "complete" model in which the spectrophotometric properties are consistently derived from the CDM linear fluctuation spectrum. Most of the objects seen in faint galaxy counts would be the low–surface–brightness dwarfs predicted by the model.

1. Introduction

In the models (or simulations) of galaxy formation which are currently developed, the mass distribution and the dynamical properties of the galaxies are computed from the linear fluctuation spectrum. But the observations give luminosity functions and colours of galaxies. So the predictions are not directly comparable with the observations, unless some mass-to-light ratio is arbitrarily fixed. Nevertheless, this procedure is far too crude, as we shall see in this paper. In order to compare the predictions of these models with observations, it is necessary to use the link of realistic models of spectrophotometric evolution. These models give meaning mass-to-light ratios and colour distributions from the small number of parameters which describe the star formation. The star formation parameters are to be consistently related to the dynamical properties of the galaxies.

The steps of a "complete" model of galaxy formation and evolution would be the following ones: (1) The particle physics give the power spectrum of the linear fluctuation field (baryonic matter + possible non-baryonic dark matter). (2) After assuming the values of H_0 and Ω_0, the distribution of collapsed haloes $n(M, z_{coll})$ (in Mpc^{-3}) is computed as a function of total halo mass and redshift of dynamical collapse, providing a Press–Schechter–like (PS) prescription, and a collapse model such as the simple top–hat model, are given. The usual PS prescription (Press and Schechter, 1974) would a priori be an oversimplification of reality. Nevertheless, Efstathiou et al. (1988a) and Efstathiou and Rees (1988) found the abundance of haloes resulting from the simulations to be in surprising agreement with the PS prediction. (3) The dynamical parameters of the haloes are easily computed from the collapse model. Then, if a baryonic density Ω_b is assumed and if the star formation processes are related to the dynamical parameters of the haloes (or of the surroundings), the model gives the distribution of galaxies $n(M_b, z_{for}, \tau_*)$ as a function of baryonic mass, redshift of first star formation and characteristic time scale for star formation. (4) By means of a model of spectrophotometric evolution, consistent mass-to-light ratios and colours are computed at each time step from the star formation history and the ages of galaxies. Finally, the model gives the observational luminosity functions $n(L, z)$ and colour distributions $n(c, z)$ at each redshift.

In fact such a program is very ambitious and a number of steps are still unclear. The purpose of this paper is to describe step (4) more extensively. Then we propose preliminary results for the "complete" program from steps (1) to (4), as resulting from a collaboration with Cedric Lacey and Joe Silk. In section 2, we describe the principles of our model of spectrophotometric evolution, referring the reader to the published papers for the technical details. In section 3, we focus on the faint galaxy counts which strongly constrain the evolution of high–redshift galaxies and the value of the deceleration parameter q_0. The natural interpretation of the data with normal luminosity evolution gives $\Omega_0 = 2q_0 \simeq 0.10$, in strong disagreement with the value of the inflationary paradigm. Some solutions to save $\Omega_0 = 2q_0 = 1$ are explored. Finally, in section 4, we describe an attempt to build a model which computes consistently the mass, light and colour distributions of galaxies at all redshifts, within an $\Omega_0 = 1$ universe filled with cold dark matter (CDM). Section 5 is our conclusion.

2. A model of spectrophotometric evolution

2.1 Principles

The basic idea of models of spectrophotometric evolution is to compute evolving synthetic spectra of the galaxies from their star formation history (Star Formation Rate and Initial Mass Function). After Beatrice Tinsley's pioneering works, the first model of this type was proposed by Bruzual (1981). We built a new model at the IAP, which is described in Rocca–Volmerange et al. (1981) and Guiderdoni and Rocca–Volmerange (1987). The latter reference gives the comparison with Bruzual's model. More recently, Yoshii and Takahara (1988) proposed predictions at high redshift, unfortunately with too crude UV stellar data.

The principles of our model are the following ones: at each time step, stars form from the gas content according to the classical parameters (SFR, IMF). These stars are placed in the Zero Age Main Sequence of the HR diagram. The model uses a compilation of stellar evolutionary tracks taking into account the Main Sequence and late stages of stellar evolution to compute at each time step the distribution of the stellar populations in the HR diagram. This distribution is combined with a library of stellar spectra in the UV and visible (220 to 10000 Å) in order to estimate the synthetic spectrum of the stellar population (Rocca–Volmerange and Guiderdoni, 1988). The internal extinction is estimated from the gas content and simple models of metal enrichment and transfer. The nebular emission is computed from the flux of ionizing photons. The mass–to–light ratios, apparent magnitudes and apparent colours at redshift z are derived from the spectra after coupling with the standard cosmological models and convolution with the response curves of filters.

2.2 Constraints from nearby and high–redshift galaxies

As shown in Table 1, the synthetic spectra satisfactorily reproduce the colours of the Hubble sequence of spectral types, with a single parameter, namely the time scale for star formation. The age of galaxies can be a secondary parameter. A universal IMF is used for all galaxies, with the slope for massive stars derived by Scalo (1986). The second important result is the large range of mass–to–light ratios predicted by the model, and shown in figure 1a. The ratio of baryonic mass to blue light varies by a factor of 30 if SFR time scales and ages are taken into account. Thus a single, arbitrary mass–to–light ratio is far from reality. If we assume that 50 % of the baryonic gas becomes Jupiters and brown dwarfs, the model for "old" galaxies (say older than 10 Gyr) predicts $M_b/L_B \sim 4$ for Im and late spirals and up to ~ 15 for E, in agreement with the observations in the central parts of galaxies.

Consistent models for the formation and evolution of galaxy are now strongly necessary to interpret the large amount of recent data on high–redshift galaxies. The strategy of search of high–redshift galaxies mainly developed into two axes: (i) *Optical and near-infrared counterparts of powerful radiogalaxies* such as the 3C, 1 Jy, 4C or Parkes catalogues. The most distant galaxies which are up to now observed have been detected by this strategy: 3C257 at $z = 2.48$, 0902+34 at $z = 3.4$ and 4C41.17 at $z = 3.8$. These objects show a number of interesting properties. The most surprising features are the strong Ly α emission in an extended component, which would imply very strong star formation rates if the ionizing photons originate in OB stars, and the alignment of the optical elongated shape with the axis

type	SF law	$B - V$ at 12–16 Gyr	$< B - V >$ in RC2
(1 Gyr burst)	$\tau(t) = 1$ $(t < 1$ Gyr$)$	1.00–0.95	
UV–hot E/S0	$1g(t)$	0.92–0.91	E: 0.89
Sa	$0.4g(t)$	0.74–0.84	0.72
Sb	$0.3g(t)$	0.67–0.78	0.66
Sc	$0.1g(t)$	0.51–0.58	0.51
Sd	0.060	0.42–0.47	0.44
Im	$7.7 \ 10^{-4} t^2$	0.31–0.34	0.36

$g(t) \equiv \mathcal{M}_{gas}(t)/\mathcal{M}_{tot}$.

Table 1: Standard scenarios of evolution for the Hubble sequence of spectral types.
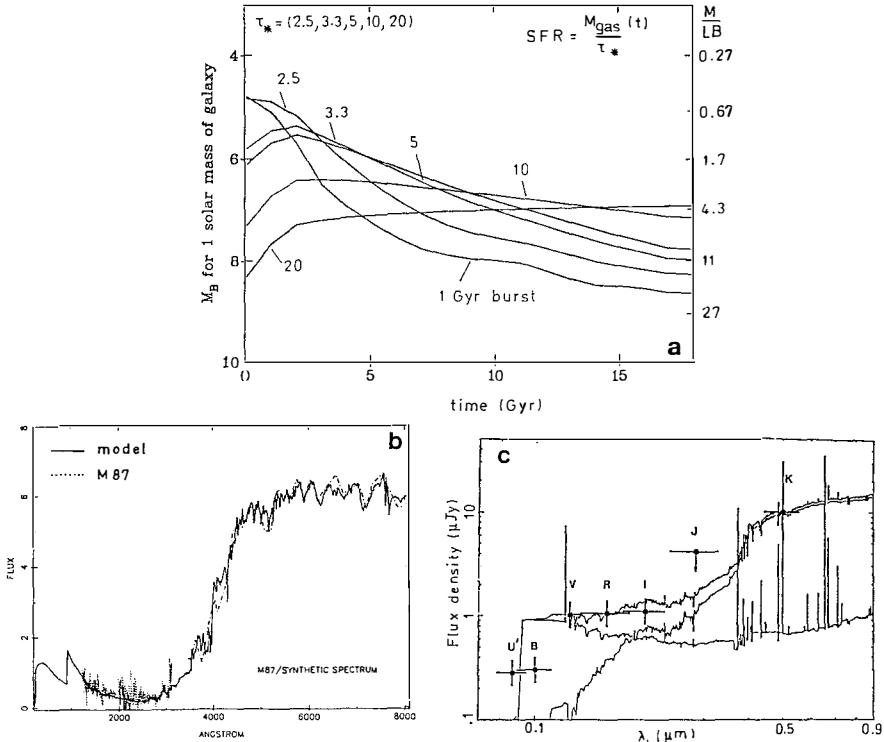

Figure 1: (a) (Baryonic)–mass–to–(blue)–light ratio against time for various star formation histories. If the various time scales t_* (in Gyr) for star formation and the age effect are taken into account, the M_b/L_B can vary by a factor ~ 30. (b) Fit of the spectrum of M87 from Rocca–Volmerange (1989) with the UV–hot model. (c) Fit of the spectrophotometry of 0902+34 discovered by Lilly (1988) at $z = 3.395$, from Rocca–Volmerange and Guiderdoni (1990) with the sum of two 0.1 Gyr bursts. The derived age is 0.5 Gyr.

of the radio jet. As shown in figure 1bc, the UV–hot model gives a good fit of the UV/visible spectrum of the nucleus of M87 (Rocca–Volmerange, 1989) and the spectrophotometry of the high–redshift radiogalaxy 0902+34 given by Lilly (1988) is fitted by Rocca–Volmerange and Guiderdoni (1990) with two 0.1 Gyr bursts of star formation. (ii) *Faint samples of field galaxies.* Figure 2ab shows that the colours predicted from the standard scenarios of evolution for the various spectral types bracket the observational colours of a "blind" deep sample of field galaxies. As shown in Guiderdoni and Rocca–Volmerange (1988), the reddest objects are reproduced if the universe is old enough, that is if $\Omega_0 \simeq 0.10$ and $H_0 = 50$. The colours of cluster galaxies at $z \sim 0.5$ are fitted with the same trend. Alternatively, it might mean that some galaxies have a high metallicity, or are strongly extincted, or that there is some problem with the time scales in the theoretical stellar tracks. Finally the deep CCD counts by Tyson (1988) provide the best constraints on the properties of the global population of galaxies at high redshift. Tyson's counts show a large number of faint blue objects, which might be young or primeval galaxies at large redshifts ($1 < z < 4$).

3. The faint galaxy counts

3.1 Standard interpretation

Among the various data sets at large z, the faint counts probe the global field population and are likely to give constraints on the average evolution of the bulk of the galaxies. The counts are weakly sensitive to the value of the Hubble constant, but at magnitudes fainter than 24, they are very sensitive to the value of q_0. Guiderdoni and Rocca–Volmerange (1990) proposed an analysis of the available data on the magnitude, colour and redshift distributions of faint galaxies. The straightforward interpretation is based on the nearby luminosity functions for each morphological type, and on the extrapolation in the past of the standard scenarios reproducing the properties of the nearby galaxies, under the assumption of pure luminosity and colour evolution. Figure 2cd shows the sensitivity of the standard predictions to q_0 and z_{for}. Thus the fit of the data in the number counts, colour histograms and redshift distribution of the Durham Faint Survey (Broadhurst et al., 1988) absolutely requires a low q_0 (~ 0.05) and a high value of z_{for} (~ 10).

3.2 Alternative interpretations

It is not easy to get rid of this conclusion. As a matter a fact, it seems to be unsensitive to uncertainties in the inputs of the model, since Yoshii and Peterson (1990) give the same conclusion on the basis of their model of photometric evolution which has other stellar data. Guiderdoni and Rocca–Volmerange (1990) showed that none of the incertainties due to the history of star formation (providing there is only luminosity and colour evolution), the IMF, the internal extinction or the nebular component can make the results consistent with the value $\Omega_0 = 2q_0 = 1$ required by the inflationary paradigm. This conclusion seems to be very robust since it is due to the smaller volume elements at high redshifts for $\Omega_0 = 1$. There are simply not enough objects in the $\Omega_0 = 1$ cosmological model to reproduce the data, whatever the luminosity of these galaxies may be.

Thus if we believe in the inflationary paradigm, we must introduce a new assumption
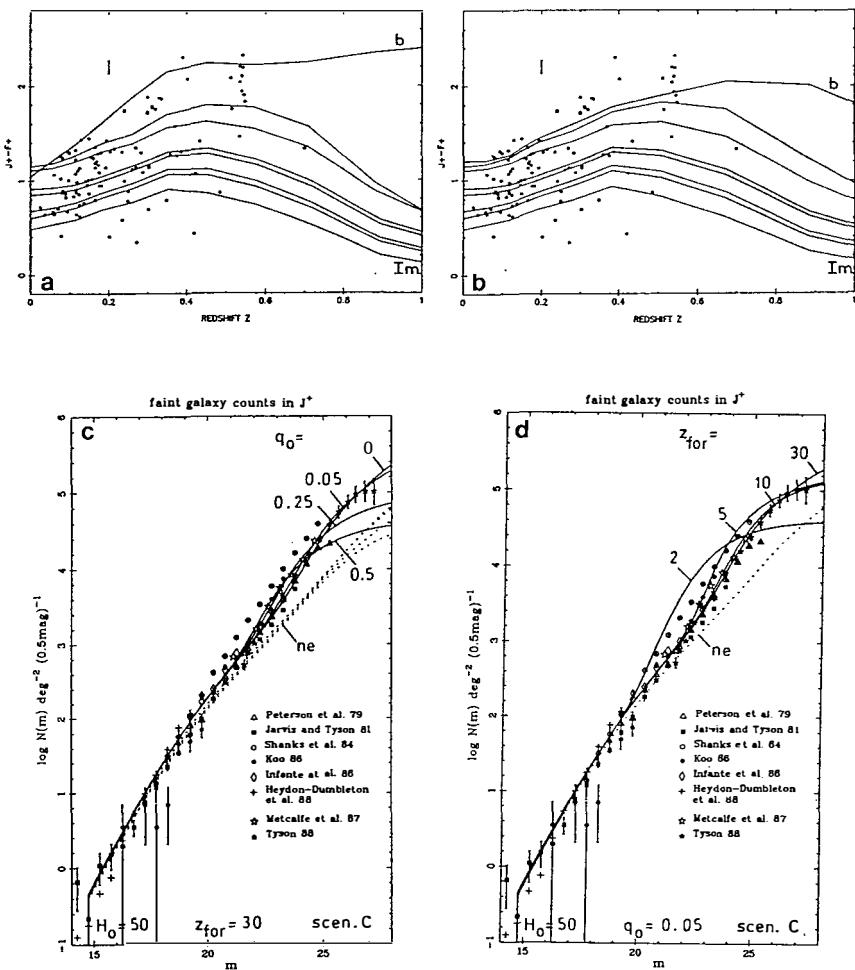
Figure 2: (a) and (b) Predicted colours for the Hubble sequence respectively with $q_0 = 0.05$ and $q_0 = 1/2$, and field sample. Reprinted from Guiderdoni and Rocca–Volmerange (1988). (c) and (d) Sensitivity of the predicted counts to q_0 and the redshift z_{for} of first star formation. Reprinted from Guiderdoni and Rocca–Volmerange (1990).

faint galaxy counts in J⁺

$H_0 = 50$ $\Omega_0 = 1$ $\eta = 2$

$\eta = 1$

$\eta = 0$

no ev.

△ Peterson et al. 79
■ Jarvis and Tyson 81
○ Shanks et al. 84
● Koo 86
◊ Infante et al. 86
+ Heydon-Dumbleton et al. 88
☆ Metcalfe et al. 87
✶ Tyson 88

$\log N(m) \deg^{-2} (0.5 \mathrm{mag})^{-1}$

m

a

redshift distribution

$H_0 = 50$ $\Omega_0 = 1$

Broadhurst et al. 88

$20 < J^+ < 21.5$

no ev. $\eta = 2$

$\eta = 1$

$\eta = 0$

$N(z)$ in gal. $(\Delta z = 0.05)^{-1}$

z

b

colour distribution

$H_0 = 50$ $\Omega_0 = 1$

$22 < J^+ < 23$

• Metcalfe et al. 87

for 0.082 deg²

$\eta = 0$

$\eta = 1$

$\eta = 2$

no ev.

$N(J^+ - F^+)$ in gal. per $\Delta(J^+ - F^+) = 0.15$ colour bin

apparent colour $J^+ - F^+$

c
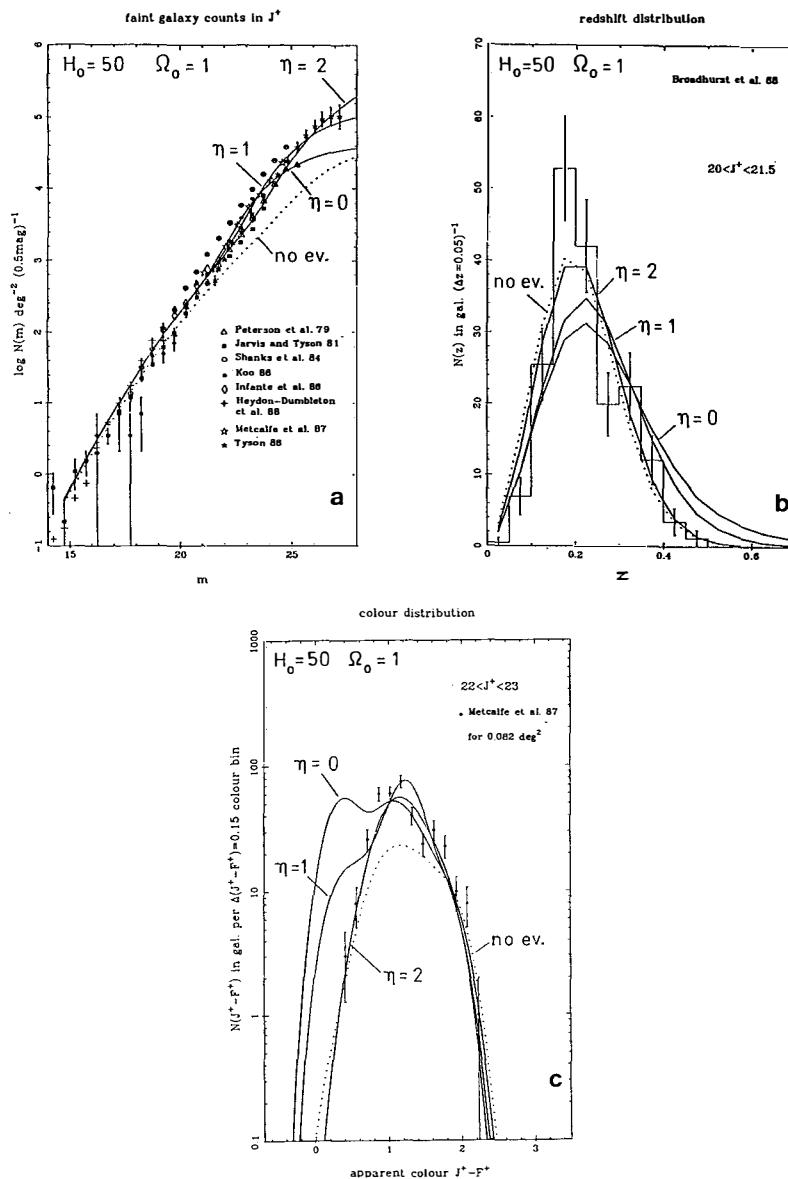
Figure 3: Predictions for number evolution as $(1 + z)^\eta$ in addition to the normal luminosity evolution. (a) Number versus apparent magnitude. (b) Redshift distribution. (c) Colour distribution.

in order to reconcile the predicted counts with the data. An appealing possibility is the introduction of a strong number evolution, in addition to the normal, unavoidable luminosity evolution. Rocca–Volmerange and Guiderdoni (1990) introduced a phenomenological modelling of the number evolution. The mass function at redshift z is the well-known Schechter function:

$$\phi(M, z)dM = \phi^*(z)\left(\frac{M}{M^*(z)}\right)^\alpha \exp{-\frac{M}{M^*(z)}} \frac{dM}{M^*(z)}$$

with $M^*(z)$ being the caracteristic mass at the knee, and $\phi^*(z)$ a caracteristic density. The values $M^*(0)$ and $\phi^*(0)$ are fixed in order to fit the current luminosities and densities of galaxies. Then $M^*(z) = M^*(0)(1 + z)^{-\eta}$ and $\phi^*(z) = \phi^*(0)(1 + z)^\eta$. The parameter η is free. This functional form has two interesting properties: (i) self–similarity, as suggested by the classical PS prescription. (ii) conservation of the total comoving mass density.

In this case, our grid of models with various laws of star formation (Rocca–Volmerange and Guiderdoni, 1988), which turns out to reproduce the observed Hubble sequence, is taken as a phenomenological description of the history of star formation in galaxies undergoing a large number of merging events: the succession of bursts of star formation associated with the merging averages into our continuous star formation rates. Finally the constraint on the average redshift of the first stellar generation is unchanged with our new tracks. We hereafter take $z_{for} = 10$. Values $z_{for} \leq 5$ are still rejected from the data, even with the number evolution. Figures 3abc respectively show the results of the models for $\Omega_0 = 1$, in the magnitude histogram down to $B_J = 27$ (a), in the redshift histogram for $20 < B_J < 21.5$ (b) and in a colour histogram (c). The value $\eta = 0$ corresponds to the exclusion of $\Omega_0 = 1$ if no number evolution is introduced. It is seen that the values $\eta \sim 1 - 2$ give fair fits of the data. The parameters $\eta = 0.5$ or 2.5 (not shown in the figures) are rejected.

Thus if galaxies evolve gently as predicted from their colours, the data tell us something about Ω_0, which has to be low. Conversely, if we believe in the inflationary paradigm, the data strongly constrain the possible evolution of galaxies. These phenomenological results are important guidelines for a subsequent modelling.

4. An attempt to build a consistent model

4.1 Principles

An alternative possibility to save the value $\Omega_0 = 1$ from the severe constraint of the faint galaxy counts is the introduction of an additional population of nearby, low–surface–brightness (LSB) dwarfs which do not appear in the existing photometric and spectroscopic survey, except in Tyson's (1988) observations. The parameters of the luminosity function can be easily selected in order to fit the data. Such a large number of dwarfs is a prediction of standard CDM models.

In collaboration with Cedric Lacey and Joe Silk, we are attempting to derive a consistent model for the formation and evolution of galaxies, from the dynamical collapse to the formation and evolution of stars. In this model, the star formation is triggered by tides due to the neighbours. The distribution of peaks in the linear fluctuation field is computed from the formalism developed by Bardeen et al. (1986), within the framework of the CDM model. The
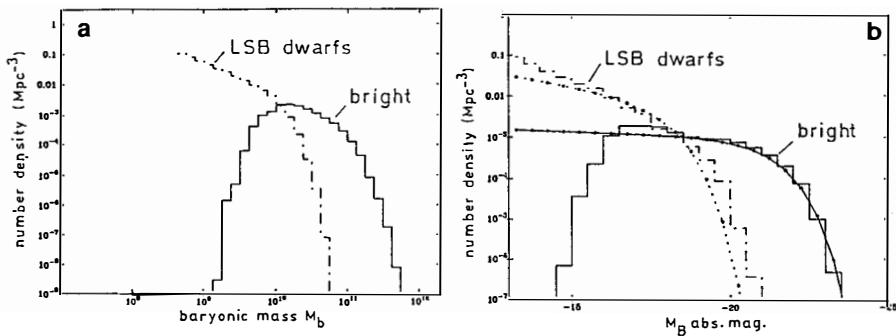
Figure 4: (a) Distribution of baryonic masses for bright galaxies (solid line) and low-surface–brightness dwarf galaxies (dashed line), computed from the model of tidally–triggered star formation given by Lacey and Silk (1990). (b) Luminosity functions computed by means of our model of spectrophotometric evolution. The data are drawn from Efstathiou et al. (1988b) and Sandage et al. (1985).
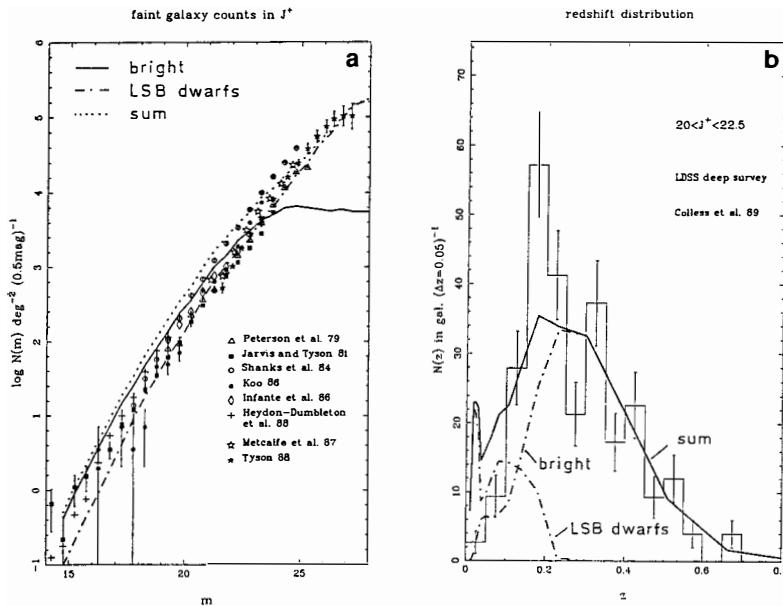

Figure 5: Predictions for the faint counts, by extrapolating in the past the luminosity functions shown in fig. 4b. (a) Number versus apparent magnitude. Most of the faint objects observed by Tyson (1988) are nearby. (b) Redshift distribution.

collapse of the peaks is modelled according to the simple top–hat model. The peak formalism also give the distribution of overdensities of the background in which galaxies are located, providing a typical size of the background with respect to the galaxies is assumed. Only those objects which can cool before the collapse of the background fluctuation are counted as galaxies. Star formation begins after the background collapse. The caracteristic time scale t^* for star formation (SFR=M_{gas}/t_*) is proportional to the collision rate in the collapsed group. The objects with too low an escape velocity loose the bulk of their gas in the galactic winds produced by supernovae, and become LSB dwarfs.

With $H_O = 50$ and $\Omega_0 = 2q_\bullet = 1$, the results are sensitive to a small number of free parameters, namely the normalisation of the power spectrum of the linear fluctuation field $\sigma(16Mpc) = 0.4$ to 0.6, the baryonic density Ω_b which is constrained by nucleosynthesis $(0.044 < \Omega_b < 0.08$ if $H_0 = 50$, from the review by Steigman, 1990) and an efficiency factor A_* relating the collision time to the star formation time scale t_*. Lacey and Silk (1990) describe the principles of the models and the first results for nearby galaxies.

4.2 Predicted luminosity functions

A typical standard model has $\sigma(16Mpc) = 0.5$, $A_* = 8$ and $\Omega_b = 0.07$. Figure 4ab shows the mass distribution of the objects and the passage to the observational luminosity function. The model satisfactorily reproduces the luminosity function of nearby, bright galaxies (Efstathiou et al., 1988b) as well as the luminosity function of LSB dwarf galaxies observed by Sandage et al. (1985) in the Virgo Cluster. The bell shape of the mass distribution is enlarged and becomes flatter because of the convolution by the wide distribution of mass–to–light ratios, giving the observed slope $\alpha = -1.07$. Figure 5ab shows the predictions for the faint counts (that is, the luminosity functions of high–redshift galaxies). Most galaxies fainter than $B_J = 22$ are in fact LSB dwarfs. The redshift distribution at $20 < B_J < 22.5$ observed by Colless et al. (1990) in the LDSS is also correctly reproduced in the model, providing a correction for isophotal magnitudes is consistently introduced into the model. Nevertheless, there is still a number of unsolved problems. In particular the predicted colour histograms for faint galaxies appear too blue, because most of the bright objects form at too low redshifts.

5. Conclusions

(1) The increasing number of data on high–redshift galaxies now actually sets strong constraints on models of galaxy formation and evolution. For interpreting these data, it is necessary to take into account properly the wide range of M/L and colours resulting from the various SFR histories, by means of a realistic model of spectrophotometric evolution.

(2) In order to save $\Omega_0 = 1$ from the faint galaxy counts, we introduce strong number evolution in the overall population of galaxies, which we constrain in a phenomenological way. Alternatively, one can introduce a population of LSB dwarfs. We describe a consistent model of galaxy formation and evolution in a $\Omega_0 = 1$ universe filled with CDM. The fit of most constraints seems to be satisfactory. Nevertheless, *all* the data have to be correctly fitted. If this turned out to be impossible, it would mean that the most natural modelling of the formation of galaxies is far too naive, or alternatively that... Ω_0 is lower than 1.

References

Bardeen, J.M., Bond, J.R., Kaiser, N., Szalay, A.S., 1986, *Astrophys. J.*, 304, 15

Broadhurst, T.J., Ellis, R.S., Shanks, T., 1988, *Month. Not. Roy. astron. Soc.*, 235, 827

Bruzual, G., 1981, Ph.D. dissertation, University of California Berkeley

Colless, M., Ellis, R.S., Taylor, K., Hook, R.N., 1990, *Month. Not. Roy. astron. Soc., in press*

de Vaucouleurs, G., de Vaucouleurs, A., Corwin, H.G., 1976, *The Second Reference Catalogue of Bright Galaxies*, University of Texas Press, Austin (RC2)

Efstathiou, G., Frenk, C.S., White, S.D.M., Davis, M., 1988a, *Month. Not. Roy. astron. Soc.*, 235, 715

Efstathiou, G., Ellis, R.S., Peterson, B.A., 1988b, *Month. Not. Roy. astron. Soc.*, 232, 431

Efstathiou, G., Rees, M.J., 1988, *Month. Not. Roy. astron. Soc.*, 230, 5p

Guiderdoni, B., Rocca–Volmerange, B., 1987, *Astron. Astrophys.*, 186, 1

Guiderdoni, B., Rocca–Volmerange, B., 1988, *Astron. Astrophys. Suppl. Series*, 74, 185

Guiderdoni, B., Rocca–Volmerange, B., 1990, *Astron. Astrophys.*, 227, 362

Lacey, C., Silk, J., 1990, *preprint*

Lilly, S., 1988, *Astrophys. J.*, 333, 161

Press, W.H., Schechter, P.L., 1974, *Astrophys. J.*, 187, 425

Rocca–Volmerange, B., Guiderdoni, B., 1988, *Astron. Astrophys. Suppl. Series*, 75, 93

Rocca–Volmerange, B., 1989, *Month. Not. Roy. astron. Soc.*, 236, 47

Rocca–Volmerange, B., Guiderdoni, B., 1990, *Month. Not. Roy. astron. Soc., in press*

Sandage, A., Binggeli, B., Tammann, G.A., 1985, *Astron. J.*, 90, 1759

Scalo, J.M., 1986, *Fundam. Cosmic Phys.*, 11, 1

Steigman, G., in *The Quest for the Fundamental Constants in Cosmology*, J. Audouze and J. Tran Thanh Van (eds.), Editions Frontières

Tyson, J.A., 1988, *A.J.*, 96, 1

Yoshii, Y., Takahara, F., 1988, *Astrophys. J.*, 326, 1

Yoshii, Y., Peterson, B.A., 1990, *preprint*

STATISTICS OF MATTER DISTRIBUTION
FROM HALO DYNAMICS

Stefano Borgani[1,2], Silvio A. Bonometto[2,3], Massimo Persic[4] and Paolo Salucci[5]

[1] S.I.S.S.A., International School for Advanced Studies, Strada Costiera 11,
    I–34014 Trieste, Italy
[2] I.N.F.N., Sezione di Perugia, via A. Pascoli, I–06112 Perugia, Italy
[3] Dipartimento di Fisica dell' Universita' di Perugia, via A. Pascoli,
    I–06112 Perugia, Italy
[4] Osservatorio Astronomico, via G.B. Tiepolo 11, I–34131 Trieste, Italy
[5] Physics Department, University of Durham, Durham DH1 3LE, England

Abstract.
By applying the dark-to-luminous mass decomposition method for spiral galaxies presented by Persic and Salucci to 58 rotation curves (with $-17.5 \geq M_B \geq -23.2$), through the deduced dynamics of dark halos we sample the galaxy–background correlation properties at short distances.

Our main results are: a) over the 20-350 h_{50}^{-1} kpc scale, the two-point galaxy–background correlation function (slope $\gamma = 1.71 \pm 0.03$; clustering length $r_{o,gb} = (6.5 \pm 0.6)(1/\Omega_o)^{1/\gamma} h_{50}^{-1} Mpc$) mimicks the two-point galaxy–galaxy function, its marginally smaller correlation length being a possible indication of bias (with a biasing parameter $b \simeq 2\Omega_o$); b) correlation functions up to sixth order fit the so-called hierarchical clustering pattern.

The main consideration which naturally follows is that on the scale of galaxies the statistics of matter seems to be highlighted with equal effectiveness by both the luminous point–object distribution and the dark–matter properties underlying the halo–influenced disk dynamics.

1. Introduction

Dark matter (DM) has long been recognized as a fundamental component of the internal structure of galaxies[1]. The overall matter distribution, as revealed dynamically by observed rotation curves, is strikingly different from the distribution of luminous matter in the optical disk region[2] and extends out to large radii where no keplerian velocity falloff is observed[3,4,5,6]. Therefore huge amounts of matter, not traced by light, dominated the overall dynamics and structure of galaxies.

In order to study the background (dark) matter statistics starting from the structure of galaxies, a general technique for decomposing the masses of spiral galaxies into their luminous and dark components is required. By applying such a technique to a sample of galaxies, the knowledge of the physical parameters of DM in the halos would permit to investigate the galaxy–background correlation properties on galaxy scales. This is particularly relevant in connection with biased scenarios for structure formation[7] which predict different though related clustering properties of matter and of cosmic structures. A suitable method, which extracts the DM content of spiral galaxies at the optical disk radius from the profiles of rotation curves, was recently devised by Persic and Salucci[8,9], and is the starting point for the present study.

Here we review and summarize the results of our two recent papers[10], in which cross–correlations between galaxy and the background distributions are investigated. Our analysis is based on the PS90 sample [8,11] of 58 non–local, Sb through Sc galaxies for which high–quality, extended rotation curves were available by the end of 1988. It spans the magnitude interval $-17.5 \geq M_B \geq -23.2$; disk radii range between 3 and 55 kpc, and peripheral velocities between 120 and 400 $km\,s^{-1}$.

Our main results can be summarized as follows:

a) the two-point *galaxy–background* correlation function has similar slope ($\gamma = 1.71 \pm 0.03$) as the two-point *galaxy–galaxy* correlation function. Its smaller clustering length, $r_c \sim 7\,h_{50}^{-1}\,Mpc$, can be interpreted as an indication of bias;

b) up to the $n = 6$, the correlation functions fit the hierarchical expression[12]

$$\xi^{(n)}(\mathbf{x}_1, \ ..., \ \mathbf{x}_n) \ = \ \sum_{n-trees} Q_{n,a} \sum_{labelings} \prod_{edges}^{(n-1)} \xi_{AB}^{(2)}; \tag{1.1}$$

c) the value of $Q \equiv Q_3$, and suitable combinations of coefficients $Q_{n,a}$ for each $n = 4$, 5, 6, are evaluated and compared with predictions of the BBGKY equation.

In view of the above results, we can state that the main statistical properties of matter on the scale of galaxies are functions of observable dynamical quantities only. Important cosmological parameters follow from observations of disk dynamics. A value of $H_o = 50\ km\ s^{-1}Mpc^{-1}$ is used throughout this paper.

2. Dark matter in the halos of galaxies.

According to Persic and Salucci[8] the disk–to–halo mass ratio in spiral galaxies at the optical disk radius, R_{25}, is directly obtainable from the profiles of optical rotation curves. By using a two–component mass model (spherical dark halo and exponential luminous thin disk) and without making assumptions on the halo density profile, we get

$$\frac{M_{disk}}{M_{tot}} = \frac{0.8 - \frac{d\log V}{d\log R}|_{R25}}{0.1 \frac{d\log V}{d\log R}|_{R25} + 1.1} . \tag{2.1}$$

Eq. (2.1) yields the disk–to–total mass ratio at R_{25} as a function of the logarithmic derivative $\frac{d\log V}{d\log R}|_{R25}$, that is obtainable from rotation curves. (Velocities V and distances R will be expressed in $km\ s^{-1}$ and kpc respectively, unless otherwise stated.)

Based on eq. (2.1), an analysis of available data[8,9] shows a number of systematic properties of dark matter (hereafter DM) in spiral galaxies: $a)$ the disk kinematics of fainter galaxies shows steeper velocity gradients, hence is mostly halo–dominated ($M_{disk}/M_{tot} \sim 0.2$), while in brighter galaxies the kinematics is mostly disk–dominated ($M_{disk}/M_{tot} \sim 0.8$). Between these extremes there is a continuous distribution of M_{disk}/M_{tot} values; and $b)$ the M_{disk}/M_{tot} distribution, owing to the correlation between galaxy luminosity and $\frac{\partial\log V}{\partial\log R}|_{R25}$, maps into the luminosity sequence of galaxies.

In order to sample the DM distribution at scales much larger than R_{25}, let us *define* an effective halo radius, R_h, as the radius encompassing an amount of DM η times greater than M_{disk}. Consequently the mean halo density at R_h is $\bar{\rho}_h = \eta M_{disk}/(\frac{4}{3}\pi R_h^3)$. Being the observed rotation curves nearly flat for $R > R_{25}$, with a value $V(R) \simeq V(R_{25}) \equiv V_{25}$[3,4], the centrifugal equilibrium condition at R_h, $M_{dyn}(R_{25}) = G^{-1}V_{25}^2 R_{25}$, implies that the volume–averaged density in the spherical halo is $\bar{\rho}_h = 4 \times 10^{-27} \frac{V_{25}^2}{R_h^2}$. The halo radius is not directly observable. However, by our definition of R_h and by eq. (2.1), it can be related to the observable quantities R_{opt} and $\frac{d\log V}{d\log R}|_{R25}$, being

$$R_h = \eta R_{25} \left(\frac{0.8 - \frac{d\ logV}{d\ logR}|_{R25}}{0.1 \frac{d\ logV}{d\ logR}|_{R25} + 1.1} \right). \tag{2.2}$$

If one takes $\eta = const$ for all galaxies, it is the variation of M_{disk}/M_{tot} with luminosity which triggers a range of R_h. In this case, if the variations of V_{25} and R_h were statistically independent, a correlation $\bar{\rho}_h \propto R_h^{-2}$ would be found. On the contrary, the observed increase of V_{25} with luminosity[2,12], owing to the above arguments on the R_h–luminosity correlation, turns into an increase of V_{25} with R_h. A rough estimate neglecting the luminosity dependence of M_{disk}/M_{tot} and using the $V_{25} - R_{25}$ correlation given in reference 2, we get

$$\frac{\bar{\rho}_h}{\rho_c} = \left(\frac{R_o}{R_h} \right)^\gamma \tag{2.3}$$

($\rho_c = 4.7 \times 10^{-30} h_{50}^2\ g\ cm^{-3}$ is the critical density) with $\gamma \simeq 1.5$ and $R_o \simeq 10\ Mpc$.
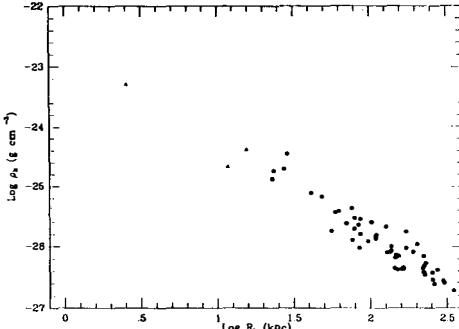
Figure .1: *The mean halo density , $\bar{\rho}_h$, as a function of the effective halo radius, R_h, for the PS90 sample.*

We assume $\eta = 10^{13,14)}$. Different choices for η reflect very mildly on R_o and only to the extent that $\gamma \neq 2$, according to $R_o \propto \eta^{\frac{\gamma-2}{\gamma}}$; no variation of γ is induced if η is chosen to be uncorrelated with luminosity.

We work out the extended halo parameters for the PS90 sample. By using eq. (2.3) we compute R_h, and then $\bar{\rho}_h$ accordingly. In Figure 1 we plot the so–computed halo densities *vs.* their respective effective radii. From these outputs we calculate the moments of mass distribution reported in Table 2 which will be used in sections 3 and 4 to work out the galaxy–background correlation functions.

3. Two–point function analysis.

We consider the *galaxy–background* correlation function $\xi_{gb}(r)$. The amount of matter within a sphere of radius R centered on a galaxy[15)] is

$$< M >_R = \rho V_R + \rho \, 4\pi \int_0^R r^2 dr \, \xi_{gb}(r) \qquad (3.1)$$

($\rho = \rho_c \Omega_o$ is the average matter density). Assuming $\xi_{gb}(r) = (r_o/r)^\gamma$, in the large-clustering ($R << r_o$) regime from eq. (3.1) we get

$$\frac{< M >_R}{\tilde{M}_R} = \xi_{gb}(R)K_1, \qquad (3.2)$$

with $\tilde{M}_R = \rho V_R$ ($V_R = 4\pi R^3/3$) and $K_1 = 3/(3 - \gamma)$.

If R is interpreted as the halo radius, by comparing eqs. (3.2) and (2.3), whereby $R_o = K_1^{1/\gamma} r_o$, the quest for a power–law ξ_{gb} could be carried out by considering $\bar{\rho}_h = M/V_R$ as a function of R for each individual galaxy (see Figure 1). However, in view of the upcoming analyses of higher–order correlations (see Sect. 4) , we collect the sample galaxies into 5 bins, each having the same logarithmic amplitude, $\kappa = 1.7358$, and a characteristic radius, $r_k = 17.0\,\kappa\,kpc$, which coincides with the logarithmic center

Table 1

Correlation length and slope of the 2-point function for different weighting in the bins.

α	$r_o \ (Mpc)$	γ
1	6.28 ± 0.58	1.71 ± 0.03
2	6.44 ± 0.60	1.71 ± 0.03
3	6.59 ± 0.62	1.71 ± 0.03

of the bin itself (see Table 2). Averaging eq. (3.2) over each bin, we obtain

$$\frac{< M >_k}{\bar{M}_k} \; = \; K_1 \int_{r_{k,min}}^{r_{k,max}} \xi_{gb}(r) p_r \, dr, \tag{3.3}$$

where $p_r \propto r^\alpha$ is a weight function such that $\int_{r_{k,min}}^{r_{k,max}} p_r \, dr \; = \; 1$.

Best–fit values of r_o and γ and standard deviations (for $\eta = 10$ and $\Omega_o = 1$) are reported in Table 1 for different values of α. We find $\gamma = 1.71 \pm 0.03$ independent of α, and $r_o \simeq (6.5 \pm 0.6) \, Mpc$ only slightly dependent of α.

Quite differently from what happens in usual correlation analysis, values coming from different bins are built with different objects hence are statistically independent.

A comparison between our outputs on ξ_{gb} and current results on the *galaxy* covariance function shows that both functions are well modeled by power laws having virtually equal slopes ($\gamma = 1.77 \pm 0.04$ for the *galaxy–galaxy* case[16)]). The respective clustering lengths, however, seem to be marginally different, being $r_{o,gb} < r_{o,gg}$. The lower value of $r_{o,gb}$ can be interpreted as the signature of a biasing in the distribution of galaxies. The amount of bias, measured by the biasing parameter $b = \xi_{gg}/\xi_{gb}$, can be evaluated by comparing $r_{o,gb}$ and $r_{o,gg}$. Then, $b = (\frac{r_{o,gg}}{r_{o,gb}})^\gamma \; \sim \; 2$, consistent with Dekel and Rees's[14)] expectation ($b \sim 2$-3), based on very general and different grounds. Due to the very mild dependence of $r_{o,gb}$ on η, our inferred value of b is hardly affected by the choice of η. We emphasize that eq. (3.2) is the first *direct* estimate of the biasing parameter b and is related to our capability of investigating the background matter statistics.

4. N–point function analysis.

In this section we test the hierarchical expression (1.1) for the n-point correlation functions, up to $n = 6$, and deduce the numerical values for suitable combinations of the coefficients $Q_{n,a}$ by means of the *moment method* (see Appendix of reference 10).

Let us consider the n–thorder moment for the mass distribution around a galaxy

$$< M^n >_R = \rho^n \int_0^R \delta V_1 ... \int_0^R \delta V_n [1 + \xi_{dis,01...n}^{(n+1)}], \tag{4.1}$$

where $\xi_{dis,01...n}^{(n+1)}$ is by definition the *disconnected* correlation function between a galaxy (chosen as the origin of the axes and labeled with zero in the following) and n points

Table 2

Central moments and population of the radial bins.

Bin limits (kpc)	$\frac{<M>_k}{\tilde{M}_k}$	$s_k^{(2)}$	$s_k^{(3)}$	$s_k^{(4)}$	$s_k^{(5)}$	Galaxies/bin
20-40	46400	17552	16740	21288	21704	4
40-70	10100	4584	2382	5109	3361	5
70-120	4350	1779	1759	2416	2627	14
120-200	1820	955	1125	1440	1604	18
200-350	808	285	272	387	413	14

lying in the background. The disconnected function contains all the lower–order correlation terms that yield the $(n - m)$-point function when m points are removed, and the *connected* terms that vanish when one point is sufficiently far away[16]. We average eq. (4.1) over each bin and obtain the quantities $< M^n >_k$, that generalize the quantity $< M >_k$, defined in eq. (3.4), to higher correlation order.

Introducing now the *central moments*

$$s_k^{(n)} = \left\langle \left[\frac{M}{\tilde{M}_k} - \frac{< M >_k}{\tilde{M}_k} \right]^n \right\rangle_k^{1/n} \tag{4.2}$$

and using the binomial expansion, it is easy to show that

$$\frac{< M^n >_k}{< M >_k^n} = \sum_{j=0}^{n} \frac{n!}{(n - j)! \, j!} \frac{(s^{(j)})^j}{(< M >_k / \tilde{M}_k)^j}. \tag{4.3}$$

Eq. (4.3) gives the connection between the observational value of $s_k^{(n)}$ (reported in Table 2) and the n-point correlation functions. Due to the large–clustering regime, in what follows we forget the non–leading contributions from disconnected terms in $\xi_{dis}^{(n)}$. Following the procedure outlined in the Appendix of reference 10, we work out suitable combinations of the hierarchical coefficients up to the sixth order:

$$Q = 0.43 \pm 0.02,$$

$$Q_{4,a} + 0.36 \, Q_{4,b} = 0.18 \pm 0.02,$$

$$Q_{5,a} + 11.43 \, Q_{5,b} + 9.90 \, Q_{5,c} = 0.67 \pm 0.11,$$

$$Q_{6,a} + 18.06 \, Q_{6,b} + 13.39 \, Q_{6,c} + 49.50 \, Q_{6,d} + 43.59 \, Q_{6,e} + 52.04 \, Q_{6,f} = 0.44 \pm 0.12. \tag{4.4}$$

According to eqs. (4.4), we detect galaxy–background correlations up to the sixth order. Data on such high–order correlations are significant as they can place constraints to the predictions of gravitational–instability evolution models. As an example, in the next section we compare the predictions of the BBGKY equation with our outputs.

Table 3
*Values of combinations of the hierarchical
coefficients, from observations and from Fry's
and Hamilton's solution of the BBGKY equation.*

Correlation order	4th	5th	6th
Observations	0.175 ± 0.015	0.675 ± 0.107	0.44 ± 0.12
Fry	0.167 ± 0.015	0.567 ± 0.081	0.70 ± 0.13
Hamilton	0.123 ± 0.011	0.787 ± 0.110	1.49 ± 0.28

5. Discussion and conclusions.

In this paper we reviewed and summarized the results of our two recent papers, in which we studied the correlation of matter with galaxy sites through the properties of dark halos of galaxies. Such properties were in turn deduced from the observed kinematics of rotating disk galaxies. This procedure establishes an operational link between galaxy structure and matter statistics which proves to be remarkably effective. Moreover, the very nature of the galaxy–background correlation function is such that even a rather limited sample of objects is sufficient to obtain a statistical significance. Computing the density of one halo surrounding a galaxy is effectively analogous to counting all the objects within a given distance of a galaxy. In addition, there is no intersection among the information coming from individual galaxies.

The 2-point correlation function is a power law with same slope as the 2-point *galaxy* function. A marginal discrepancy between the respective clustering lengths might be indicative of bias (with biasing parameter $b \sim 2$). The range of relevant scales widely overlaps with the range sampled through galaxy counts, our data extending to even slightly smaller scales.

As for the higher-order correlations, a significant point is that the hierarchical pattern [see eq. (1.1)] is warranted by observational data up to order $n=6$. Suitable combinations of the hierarchical coefficients $Q_{n,a}$ also are obtained. A comparison of our outputs with analogous results from the literature (by galaxy number counts techniques) is discussed in reference 10 for the 3- and 4-point functions. Our 5- and 6-point results can be tentatively compared with the very recent results based on galaxy counts[17].

The statistically significant signal for the combinations of hierarchical coefficients warrants comparison with the predictions of the two hierarchical solutions of the BBGKY equation[18] by Fry[12] and by Hamilton[19]. According to Fry, all the *trees* in eq. (1.1) have equal amplitudes $Q_{n,a} = Q_n = \left(\frac{4Q}{n}\right)^{n-2}\frac{n}{2n-2}$. On the other hand, Hamilton argues that eq. (1.1) can be taken as a solution of the BBGKY equation when only contributions from *snake* graphs are considered. Consequently $Q_{n,snake} = Q^{n-2}$ and $Q_{n,non-snake} = 0$.

Comparison between the two sets of predictions and our outputs (see Table 3) shows that Fry's solutions are in significantly better agreement. In fact that model

374

predicts the values of those combinations remarkably well up to the 5th order, and still within $\sim 1.5\sigma$ for the 6th order.

As a concluding remark, our approach to the study of matter statistics involves a different physical basis as compared to the usual analysis of galaxy clustering. The essential feature that allows us to link rotation curves and background (dark) matter distribution is our capability to extract the DM content of spiral galaxies at one same physiscal radius (in galaxy lengthscale units) for all galaxies. In our opinion, our results highlight a deep underlying connection between internal structure and large–scale distribution of galaxies.

References

1. Faber, S.M., and Gallagher, J.S. 1979, *Ann. Rev. Astr. Ap.*, **17**, 135.
2. Rubin, V.C., Burstein, D., Ford, W.K., Jr., and Thonnard, N. 1985, *Ap.J.*, **289**, 81.
3. Bosma, A. 1981, *A.J.*, **86**, 1825.
4. Begeman, K. 1988, *Ph.D.* thesis, Groningen.
5. Carignan, C., and Freeman, K.C. 1988, *Ap.J. (Letters)*, **332**, *L33*.
6. Carignan, C., Sancisi, R., and van Albada, T.S. 1988, *A.J.*, **95**, 37.
7. Kaiser, N. 1984, *Ap.J. (Letters)*, **284**, L9;
   Jensen, L.G., and Szalay, A.S. 1986, *Ap.J. (Letters)*, **305**, L5;
   Politzer, D., and Wise, M. 1984, *Ap.J. (Letters)*, **285**, L1;
   Bardeen, J.M., Bond, J.R., Kaiser, N., and Szalay, A.S. 1986, *Ap.J.*,**304**, 15.
8. Persic, M., and Salucci, P. 1988, *M.N.R.A.S.*, **234**, 131.
9. Persic, M., and Salucci, P. 1990, *M.N.R.A.S.*, in press.
10. Bonometto, S.A., Borgani, S., Persic, M., and Salucci, P. 1990, *Ap.J.*, in press;
    Borgani, S., Bonometto, S.A., Persic, M., and Salucci, P. 1990, *Ap.J.*, submitted.
11. Persic, M., and Salucci, P. 1990, *Ap.J.*, in press.
12. Fry, J.N. 1982, *Ap.J.*, **262**, 424; Fry, J.N. 1984, *Ap.J. (Letters)*, **277**, L5.
13. Tully, R.B., and Fisher, J.R. 1977, *Astr. Ap.*, **54**, 661;
    Richter, O.-G., and Hutchmeier, W.K. 1984, *Astr. Ap.*, **132**, 253;
    Aaronson, M., and Mould, J. 1986, *Ap.J.*, **303**, 1.
14. Blumenthal, G.R., Faber, S.M., Primack, J.R., and Rees, M.J. 1984, *Nature*, **311**, 517.
15. Dekel, A., and Rees, M.J. 1987, *Nature*, **326**, 455.
16. Peebles, P.J.E. 1986, *Nature*, **321**, 27.
17. Szapudi, I., Szalay, A.S., and Boschan, P. 1989, *P.A.S.P.*, in press.
18. Davis, M., and Peebles, P.J.E. 1977, *Ap.J. Suppl.*, **34**, 425.
19. Hamilton, A.J.S. 1988, *Ap.J.*, **332**, 67.

COSMOLOGICAL MASS FUNCTIONS: THEORY AND APPLICATIONS

J. A. Peacock

Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ.

Abstract. A review is given of existing analytic methods for calculating the mass distribution of bound objects which form in an expanding universe. Special emphasis is given to results for hierarchical theories, and especially to methods applicable to Gaussian density fields. Improvements and modifications to the Press-Schechter formalism are surveyed; the fact that this method works so well in practice seems to be a happy accident. Some of the pitfalls to be negotiated in making the transition from collisionless halo multiplicity functions to luminosity functions for galaxies and clusters are described, with particular reference to biased galaxy formation. Applications of these methods abound; among the most important are the abundance and evolution of rich clusters, the interpretation of galaxy redshift survey data and the quasar redshift cutoff.

376

1 Introduction

The luminosity functions which are important in cosmology have long been known to display an intriguing simplicity. Over perhaps two decades of luminosity, the distribution functions for both galaxies and groups/clusters can be well described by the Schechter form

$$d\phi(L) = \phi^* \left(\frac{L}{L^*} \right)^{-\alpha} e^{-L/L^*} \cdot \frac{dL}{L}, \tag{1}$$

where $\alpha \simeq 0.1$ for galaxies (Efstathiou *et al.* 1988a) and $\alpha \simeq 1$ for groups/clusters (Bahcall 1979). Any cosmologist who fails to be challenged by this simplicity clearly lacks a soul; unless we can explain why α and L^* take the values they do, we clearly do not understand very much about galaxy formation. If we *can* explain what we see, however, the possibility opens up of using these luminosity functions as a probe of the conditions which gave rise to structure formation in the universe.

There are a number of choices to be made before we can proceed, given the large number of ways in which structure formation may take place. The most fundamental is whether to work with the gravitational instability picture, in which small initial density perturbations are amplified, or to opt for a picture where already-nonlinear perturbations are produced, for example by topological defects. This review will concentrate on the former approach, not only because it requires fewer assumptions, but because it is easier to do calculations leading to testable predictions (for a review which takes the opposite approach, see Brandenberger 1990). Furthermore, we shall concentrate on density fields where structure forms 'bottom up' or hierarchically. In large-scale damping pictures such as pure baryon or neutrino-dominated adiabatic universes, all structures smaller than superclusters form by dissipative fragmentation, so the problem is more difficult than with a hierarchy. Although facts such as the relatively small velocity perturbations in the local supercluster plus constraints on ages of high-redshift galaxies may argue that galaxies do form first, we should recognise that the main reason so much work has been done in this area may be just that people tend to pursue the theoretical line of least resistance.

Lastly, some terminology. The comoving number density of objects in the mass range dM will be called the *mass function*, and denoted by $f(M)$. The product $M^2 f(M)$ therefore gives the contribution to the total density of objects lying in unit range of $\ln M$, and is usually known as the *multiplicity function*.

2 Non-linear methods

The problem we are dealing with is inherently a non-linear one, and there have been various attempts to meet this head-on by seeking a self-consistent description of the mass distribution which encompasses both the multiplicity function and the clustering properties of the matter. These take the point of view that the hierarchy of n-point correlation functions provide

a complete description of the statistics of the density field, and so should be capable of generating the multiplicity function.

The approach of Balian & Schaeffer (1989) is empirical, relying on the power-law behaviour of the two-point correlation function $\xi(r) \propto r^{-\gamma}$ plus the assumption that the n-point function may be factorised:

$$\xi_n(r_1, \cdots, r_n) = Q_n \sum \prod_{}^{n-1} \xi(r_{ij}).$$ (2)

Given an assumed scaling law for the Q_n, the conditional probability of having N objects within a volume V, $P(N|V)$, may be calculated. Balian & Schaeffer (following Schaeffer 1987) convert this into a luminosity function, although the details and assumptions which go into this calculation are not yet published. This yields an expression for the galaxy luminosity function in quite good agreement with observation. Expressed as a mass function (constant M/L), one case of their expression has the approximate form

$$f(M) \propto \left(\frac{M_*}{M}\right) \left(\ln \frac{M_*}{M}\right)^2 e^{-M/M_*}.$$ (3)

A more direct route in this line is provided by Saslaw & Hamilton (1984), who argue thermodynamically to reach

$$P(N|V) = e^{-\bar{N}(1-b)-Nb} \frac{\bar{N}(1-b)}{N!} [\bar{N}(1-b)+Nb]^{N-1},$$ (4)

where $\bar{N}$ is the mean density and b is a parameter describing the ratio of the correlation potential and kinetic energies. However, it has to be said that this approach has generated some controversy, and there is at present no real consensus on its applicability.

A more radical approach has been taken by Henriksen & Lachièze-Rey (1990). They argue on the basis of turbulence theory that a certain spectrum of structure is to be expected, irrespective of the origins of the perturbations. In their notation, the two-point correlation function is $\xi(r) \propto r^{-D_\rho}$, and the mass function is predicted to be

$$f(M) \propto M^{-3+2/(3-D_\rho)} \exp\left[-[M/M_*]^{D_\rho/(6-2D_\rho)}\right].$$ (5)

There are clearly some successes here, in that mass functions of the required form (power law with quasi-exponential cutoff) are indeed generated. However, there are also some serious problems, which can leave one wondering how seriously to take the results. In particular, these methods are mostly restricted to self-similar distributions, and there is no indication of how to proceed if $\xi(r)$ is not exactly a power law. Also, there is no way of relating the clustering hierarchy to any linear initial conditions; critical parameters such as M_* cannot be calculated from first principles, and so there is a lack of predictive power. This is a serious defect if we are interested in testing inflationary models for the origin of fluctuations.

3 Linear-based methods

Having seen the restrictions of fully non-linear methods, we come to the area which has been pursued most energetically in recent years: using modifications of linear theory to tackle non-linear situations. Although this may sound a contradiction in terms, there are a number of approaches which can yield useful results.

3.1 N-body simulations

It may seem unconventional to include these under the heading of linear theory, but such simulations do provide the simplest method of calculating the forward evolution of a mass distribution from linear initial conditions. During much of the 1980s, N-body simulations were used as a primary tool to investigate the applicability of galaxy formation models such as Cold Dark Matter (White *et al.* 1987a and refs therein). Although exact in principle, they do have their limitations – mainly in lack of resolution. The largest simulations consider $\sim 10^7$ particles, but this only allows a dynamic range in mass of $\lesssim 100 - 300$ in determining the multiplicity function in a manner free from numerical artifacts (Efstathiou *et al.* 1988b).

More significantly, attention has now shifted towards simulations which follow the dissipative processes which occur after non-linear collapse of a collisionless dark matter halo (*e.g.* Carlberg & Couchman 1989, Katz & Hernquist 1989). Since no simulation can resolve the shocks which are actually responsible for the dissipation, nor follow any subsequent star formation, these processes are inevitably put in 'by hand' in a rather ad hoc manner. In this case, one may as well abandon any pretence of exactness, and treat the gravitational dynamics in an approximate analytic fashion also.

3.2 Burgers' equation

Also known as the *Adhesion model*, this is a method which effectively modifies the well-known approximate non-linear theory of Zeldovich (1970) to deal with the motion of sticky particles. The use of Burgers' equation in connection with the problem of large scale structure has been discussed in several papers (Gurbatov, Saichev & Shandarin 1985, 1989; Shandarin, 1987). For a review, see Shandarin & Zeldovich 1989.

The Zeldovich approximation attempts to solve the pressure-free equation of motion by extrapolating the initial linear displacements:

$$x_i = q_i + b(t)s_i(\mathbf{q}), \tag{6}$$

where x_i and q_i are co-moving Eulerian and Lagrangian coordinates respectively, $b(t)$ is a function describing the time evolution of the growing mode of gravitational instability in the linear approximation and $s_i(\mathbf{q})$ is the potential vector field

$$s_i(\mathbf{q}) = \frac{\partial \Phi_0(\mathbf{q})}{\partial q_i}, \tag{7}$$

where Φ_0 is proportional to the initial gravitational potential:

$$\phi = 3\bar{a}ab\Phi_0(\mathbf{q}). \tag{8}$$

Now, by defining a new set of variables to replace peculiar velocity (V_i) and density (ρ)

$$v_i(x_i, t) = \frac{1}{ab}V_i(x_i, t) \tag{9}$$

$$\eta(x_i, t) = a^3\rho(x_i, t) \tag{10}$$

and applying the Zeldovich approximation, the equations of motion become those of a perfect force-free fluid.

The Zeldovich approximation is exact in one dimension and works well in 3D initially, failing once trajectories cross (*e.g.* Efstathiou 1990). Gurbatov, Saichev & Shandarin realised that the addition of an infinitesimal viscosity term to the perfect fluid equations would cause these intersecting particles to 'stick', allowing the approximation to be continued to later time. The equation then becomes a standard one in fluid mechanics (Burgers 1940, 1974) and the solution may be found geometrically, as follows. It should be obvious that the Zeldovich approximation corresponds to finding the minimum with respect to $\mathbf{q}$ in the following function:

$$G(\mathbf{x}, \mathbf{q}, b) = \frac{(\mathbf{x} - \mathbf{q})^2}{2b} - \Phi_0(\mathbf{q}), \tag{11}$$

which has the geometrical interpretation that one finds the paraboloid which is tangent to the potential function (hyper)surface at the Lagrangian point $\mathbf{q}$, the Eulerian position being at the minimum of the parabola. The adhesion model modifies this by saying that, if the tangent parabola intersects the potential surface elsewhere, then the material at $\mathbf{q}$ has been bound into a 'pancake'; the edges of the pancakes are delimited by finding the parabolae which are tangent at two or more points.

The advantage of the adhesion model is that it gives a solution at arbitrary time without the need for iteration. Efficient algorithms can find the pancakes in a Burgers simulation in a time of the same order as is needed to compute the gravitational forces in an N-body code ($\sim N \ln N$), so Burgers wins by the number of time steps (usually $\sim 10^2$). Algorithms for implementing this approach in 3D have been developed by Weinberg & Gunn (1990). Working in one dimension has some advantages, however, both because it is easier to locate the pancakes and because of the exactness of the method in this case. Here, one is simulating a set of self-gravitating sheets, and the force on a particle at a given time is independent of distance from a sheet; this leads to the Zeldovich ansatz in the absence of sheet crossings. Although less relevant to the real world, the one-dimensional adhesion model can serve as a convenient test-bed for analytical approximations. A series of such one-dimensional simulations has been carried out by Williams *et al.* (1990; WHPS) and we shall discuss some of the results from this work in the next section.

4 Press-Schechter and relations

The Burgers' equation approach relies on using the linear gravitational potential; is it possible to make progress using the linear *density* field itself? This is the basis of the method introduced by Press & Schechter (1974; PS), in what has proved a most influential paper. The amount of interest in this method merits devoting a complete section to it. Most applications have been made to density fields with random phases and hence Gaussian statistics. These fields are the most natural outcome of quantum fluctuations during inflation, as well as being easier to analyse. The following discussion will therefore concentrate mainly on this case; see Lucchin & Matarrese (1988) for a detailed discussion of the application of the PS method to non-Gaussian fields.

4.1 Fundamentals

Consider some density field which is allowed to grow to non-linearity, where the initial power spectrum contains contributions at two main wavelengths, the long wavelength term having lower amplitude. Initially, non-linear clumps of size $\sim \lambda_{\min}$ will form, but with a number density modulated on the scale $\lambda_{\max}$. What is the relation of this modulation to the initial linear large-wavelength disturbance? The PS assumption is that by *filtering* the non-linear density field until $\delta\rho/\rho \ll 1$, one recovers a large-wavelength disturbance which will grow according to linear theory. This seems intuitively reasonable, at least for density fields containing enough large-scale power; we shall see below just how much is needed. The functional form of the filter is arbitrary, and different forms will give different results in detail. Before pursuing this point, we first go briefly over the original PS argument.

PS's main assumption was that a region of space which on average had $\delta\rho/\rho$ (henceforth simply δ) of order unity would collapse to a bound object. For example, in spherical symmetry the critical value according to linear theory with $\Omega = 1$ is $\delta_c = 1.686$. So, assuming we know the statistics of the linear density field, we can filter the field with a function posessing some scale size R_f and ask how the fraction of points with $\delta > \delta_c$ varies with R_f (*i.e.* with mass, assigning a mass $M \propto R_f^3$). If $\delta > \delta_c$ is assumed to assign that point to a system with mass $> M$, then this integral probability is related to the mass function $f(M)$ via

$$Mf(M)/\rho_0 = |dp/dM|, \qquad (12)$$

where ρ_0 is the total comoving density.

The main problem with this method arises immediately: the resulting mass function is not normalised. Even for $R_f \to 0$, most random fields will be symmetric between over- and under-dense regions, so that $p(\delta > \delta_c) \to 1/2$. PS decided to renormalise, by dividing by $p(\delta > 0)$ – *i.e.* multiplying by a factor 2 in the Gaussian case. For a Gaussian field, the statistics are specified by just the variance in δ: $\sigma^2(M)$, and PS deduced

$$f(M) = \frac{2\delta_c}{\sqrt{2\pi}\,\sigma} \frac{\rho_0}{M^2} \left| \frac{d\ln\sigma}{d\ln M} \right| \exp(-\tfrac{1}{2}\delta_c^2/\sigma^2). \qquad (13)$$

For spectra of power-law form, which we will often wish to consider as an illustration,

$$|\delta_k|^2 \propto k^n \tag{14}$$

the variance is $\sigma^2 = (M_*/M)^{(n+d)/d}$ (in d dimensions) and the PS mass function takes the form

$$f(M) \propto M^{-2+(n+d)/2d} \exp\left[-\tfrac{1}{2}(M/M_*)^{(n+d)/d}\right]. \tag{15}$$

Apart from the notorious normalisation problem, there is also the question of connectivity. If the average density within a region of size $\sim R_f$ has $\delta > \delta_c$ then that whole region collapses. However, to achieve this it is necessary only to have the filtered $\delta > \delta_c$ *at one point*. The PS prescription may then assign that point to a high-mass object, even though it is surrounded by points assigned to objects of lower mass. Given these difficulties, it is understandable that the PS method languished in partial neglect until it began to be clear that, whatever the problems of principle, the method works! Figure 1 illustrates the excellent fit to N-body data.

Notice that the shape of the multiplicity function is rather sensitive to the power spectrum, being broader for those spectra with more large-scale power (negative n). Although there is not enough dynamic range in mass to test very well the predicted power-law behaviour of $f(M)$ at low masses, it is clear that the PS formula follows the variations between the different power spectra rather well. Certainly, it will be a challenge for any other prescription to do as good a job. As an example, consider the formula of Henriksen & Lachièze-Rey. Equality of low-mass slope with PS requires $D_\rho = (3n + 15)/(n + 9)$, which would require much flatter correlation functions than are seen in the simulations.

4.2 Limits on spectra

The assumptions that go into PS type models are only valid for a limited range of spectral indices. It had been previously thought by various authors (*e.g.* PS; Peebles 1980) that the growth of non-linear perturbations would vary with scale factor as

$$M_* \propto a^{2d/(n+d)} \tag{16}$$

(where d is the number of dimensions) provided that the spectrum was flatter than the 'minimal' value $n = 4$ which results from rearranging uniform mass in a momentum-conserving manner. For $n > 4$, it was proposed that the growth would take a limiting value given by

$$M_* \propto a^{2d/(4+d)} \tag{17}$$

owing to non-linear generation of the long wavelength part of the spectrum where mode-coupling terms dominate over any intrinsic large scale power. However, it was suggested by Gurbatov, Saichev & Shandarin (1989) that the growth rate would only be independent of n
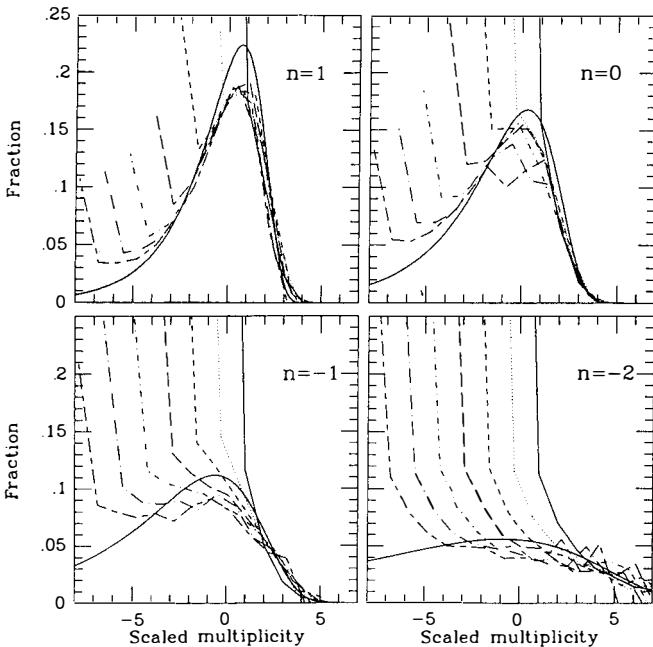
Figure 1. *The multiplicity function data of nonlinear clumps from the hierarchical simulations of Efstathiou et al. (1988b). The different epochs are scaled in mass and compared to the PS formula using a spherical top-hat filter. The fit is good, especially when it is considered that the mass scaling is not forced to provide a best fit. However, note that there is a tendency for PS to over-predict the numbers of objects around the peak of the multiplicity functions*

while the variance of the gravitational potential, $\sigma_{\Phi_0}^2$, remains finite – *i.e.* for $n > 1$ in three dimensions and for $n > 3$ in one dimension, assuming a cutoff at small wavelengths.

This conclusion is backed up by the work of Kida (1979) who effectively (although working in the context of turbulence theory) solved the Burgers model exactly in one dimension to obtain $f(M)$ in the limit of very steep spectra. Not only does M_* scale as $a^{1/2}$ rather than $a^{2/5}$, but the form of $f(M)$ is greatly different from the PS prediction. WHPS's simulations for spectral indices above $n = 3$ show that this is indeed the point where the PS mass function breaks down in 1D. This supports previous work on one dimensional power spectra by Kotov & Shandarin (1989).

Figure 2 shows a plot of $M^2 f(M)$ for $n = 1$ and 10, together with PS and Kida's
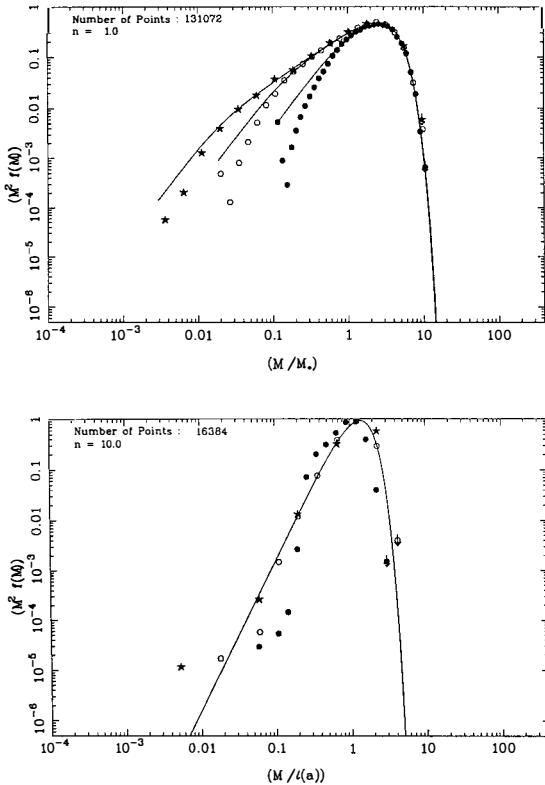
Figure 2. *The pancake multiplicity function for a Burgers' equation simulation in one dimension, with (a) an $n = 1$ spectrum; (b) $n = 10$. For $n = 1$, the fit to PS is very good; for $n = 10$ the fit is the theoretical curve for the limit of very steep spectra from Kida (1979). The different symbols refer to different epochs.*

$f(M)$:

$$f(M) = \frac{\pi}{2} \frac{M}{\ell(a)^3} \exp\left[-\frac{\pi}{2}\left(\frac{M}{\ell(a)}\right)^2\right] \tag{18}$$

where the cutoff $\ell(a)$ grows as $a^{1/2}$. It is interesting to compare this to the PS mass function under the assumption that for $n \geq 3$ in one dimension, one may use a constant value of $n = 3$ in the mass function. Both mass functions give the correct low mass slope ($f(M) \sim M$ for $M \ll M_*$, where $M_* \sim \ell(a) \sim a^{\frac{1}{2}}$) but the cut-offs are different, Kida's formula going as $\exp[-(M/M_*)^2]$ whereas the PS mass function would predict a cut-off going as

$\exp[-\frac{1}{2}(M/M_*)^4]$, *i.e.* PS cuts off too quickly, not producing enough high mass objects. Nevertheless, as Figure 2a illustrates, PS works even better in one dimension than in three, provided $n < 4 - d$. In this case, we do have sufficient dynamic range in mass to confirm that the low-mass portion of $f(M)$ follows very accurately the PS power law. The fact that the PS formula works so well is immensely frustrating, given the problems with its derivation. It is clearly worth some effort to try to understand how this can come about.

4.3 Random trajectories

To improve on the PS analysis, it is necessary to treat the points with $\delta < \delta_c$ explicitly. This may be tackled by considering the trajectory taken by δ at a given point as a function of R_f:

$$\delta(R_f|\mathbf{x}) = \frac{V}{(2\pi)^3} \int \delta_k \, F_k(R_f) \, e^{-i\mathbf{k}\cdot\mathbf{x}} \, d^3k, \tag{19}$$

where V is the normalisation volume, F_k is the k-space filter. In general this random trajectory will *not* be a random walk: $d\delta/dR_f$ will be correlated with δ. Now, for a set of very widely spaced R_f values, the fields will be essentially independent random samples and, if $\sigma(R_f) \to \infty$ as $R_f \to 0$ (*i.e.* we are dealing with a hierarchical density field which lacks a low-wavelength cutoff), then it is certain that δ will exceed δ_c on some small scale. The important question is then to find the *largest* filter value for which δ is equal to the threshold, known as the first upcrossing of the process $\delta(R_f)$. This finds the largest mass which has collapsed about that point by the present epoch, destroying in the process any sub-structure (the 'cloud-in-cloud' problem). Analyses which consider this trajectory viewpoint have been performed by Peacock & Heavens (1990) and Bond *et al.* (1990), and these shed considerable light on the operation of the PS model. These are, however, restricted to Gaussian fields: the application of such ideas to non-Gaussian fields is still an open problem.

The probability distribution for the upcrossing points may be constructed by writing

$$p(> R_f) = p_0(\delta > \delta_c) + \int_{-\infty}^{\delta_c} \frac{dp_0}{d\delta} \, p_{up}(\delta_c, \delta) \, d\delta, \tag{20}$$

where p_0 is the unconstrained distribution. This says we should divide points at R_f into two classes. Those with $\delta > \delta_c$ clearly have $\delta = \delta_c$ for some larger filter, and are therefore associated with objects of scale $> R_f$. For all points below the threshold, there is some probability p_{up} that subsequent filterings might at some point result in having δ above the threshold. This prescription is automatically normalised: it is clear that $p_{up} \to 1$ as we go to very small scales. The PS argument just corresponds to taking only the first term in the above equation and doubling it.

Solving exactly for the upcrossing probability in this way is hard, but a method for circumventing the difficulties is given by Peacock & Heavens (1990). Once the field has been filtered by a large amount, the result is essentially independent of the original value and the probability of exceeding the threshold is just the unconstrained value; the same is true for

all subsequent large filterings. The *survival probability* (the probability of always remaining below the threshold; $p_s \equiv 1 - p_{up}$) is then just the product over these independent fields:

$$p_s(\delta_c) = \prod_i p_i(\delta < \delta_c).$$ (21)

However, this relation cannot hold for a set of fields with very similar R_f values; there must exist the analogue of a coherence length in $\log R_f$. This suggests the following simple ansatz for the survival probability:

$$\ln p_s = \int_0^\infty \ln p(\delta < \delta_c | R_f) \, \frac{d \ln R_f}{\Delta},$$ (22)

where Δ is the critical increment in $\ln R_f$ for effective independence (at this stage, Δ is simply a parameter to be determined, and will depend on the power spectrum).

Peacock & Heavens (1990) show how Δ can be determined by looking at the set of points which have $\delta < \delta_c$, and evaluating $p(\delta > \delta_c)$ for an infinitesimal amount of smoothing. This simplifies the correlated random walk by only considering the initial step. For Gaussian filtering, $F_k = \exp(-\frac{1}{2}k^2 R_f^2)$, one obtains

$$\Delta = \frac{\pi \ln 2}{R_f^2} \left[\frac{\sigma_2^2}{\sigma_0^2} - \frac{\sigma_1^4}{\sigma_0^4} \right]^{-1/2}.$$ (23)

It is straightforward in principle to derive the corresponding expressions for other filter functions. In general, these will involve k-space integrals over derivatives with respect to R_f of the k-space filter function. Gaussian filtering is analytically convenient, in that these integrals reduce to moments of the power spectrum.

It should be immediately clear that this solution to the PS problem may in general lead to results rather different from the original PS formula. Differentiate the fundamental equation to get the differential distribution of R_f values: in the limit of small R_f, where $p_{up} \to 1$, this becomes simply

$$\frac{dp}{dR_f} = \frac{1}{2} \frac{dp_{up}}{dR_f}.$$ (24)

Since the survival probability only goes to zero as a slow power law $p_s \propto R_f^{(\ln 2)/\Delta}$, the possibility is raised that this term might tend to dominate the mass function for small R_f. We now show that this is indeed the case: this formalism will usually predict more low-mass objects than the original PS formula.

In presenting results, it is convenient to define the mass measure

$$\nu \equiv \frac{\delta_c}{\sigma(R_f)},$$ (25)

where $\nu \propto M^{(n+d)/2d}$ for power-law spectra. Since the multiplicity function is given by

$$M^2 f(M)/\rho = \left(\frac{dp(\delta > \delta_c)}{d \ln \nu} \right) \left(\frac{d \ln \nu}{d \ln M} \right),$$ (26)
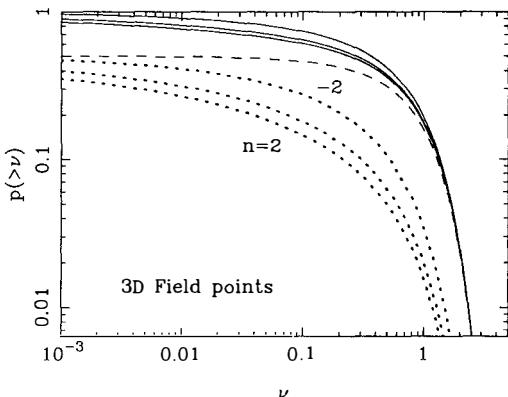
Figure 3. *The integral probability distribution for first upcrossing points of a Gaussian-filtered three-dimensional random Gaussian field. The filter radius of upcrossing is specified in terms of $\nu \equiv \delta_c/\sigma(R_f)$. Three curves are shown: the fraction of mass which lies above the threshold at a given scale (dashed); the fraction which lies below the threshold, but lies above the threshold on some larger scale (dotted); and the total (solid). The first fraction is always the larger, and corresponds to half the PS result.*

where the second term is a constant for power-law spectra, we see that the PS result may be written for all n as

$$M^2 f(M) \propto \nu e^{-\nu^2/2}. \tag{27}$$

Figure 3 shows the probability distribution for the upcrossing point in integral form.

This shows several interesting features. At large masses, $\nu \gtrsim 2$, we attain exactly the PS result without rescaling. Since we need to achieve $p = 1$ for very low masses, it is clear that the differential probability will be larger than PS at $\nu \ll 1$: the missing factor of 2 at large masses must be supplied by low-mass objects. This is shown directly in Figure 4.

Although the figure considers Gaussian filtering only, it seems likely that the behaviour will be qualitatively the same in most cases. The fact that $P(> \nu)$ goes to unscaled PS for large masses just says that, for points with $\delta < \delta_c$ at high mass, it is very improbable that filtering will increase δ: most $\delta(M)$ trajectories will fall at high M.

The exception to the above discussion comes when the first derivative of k-space filter function with respect to R_f does not exist. The main example of this case is sharp truncation in k-space. This has been investigated recently in the context of the PS model by Bond *et al.* (1990) and has some appealing properties. Crucially, the field in this case executes a random walk from $\delta = 0$ at $R_f = \infty$: each new slice of k-space added is independent. In this case, one can use the following argument to deduce the survival probability (Chandrasekhar 1943).
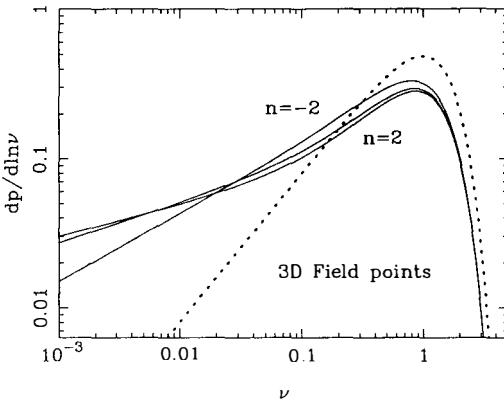
Figure 4. *The total results of Fig. 3 in differential form, $dp/d\ln\nu$. Since $M^2 f(M)/\rho_0 = (dp/d\ln\nu)(d\ln\nu/d\ln M)$, these plots are identical in shape to the multiplicity function. The PS result (which is the same for all spectra) is shown dotted.*

For $\delta < \delta_c$ the probability density of points which have upcrossed at a larger R_f must, by the symmetry of the random walk, simply be the reflection in the threshold of those points above the threshold. The upcrossing probability is then just

$$p_{up} = \frac{e^{-(\delta-2\delta_c)^2/2\sigma^2}}{e^{-\delta^2/2\sigma^2}}, \tag{28}$$

and the total probability of upcrossing is

$$\int_{-\infty}^{\delta_c} \frac{dp_0}{d\delta} \, p_{up}(\delta_c,\delta) \, d\delta = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\delta_c} e^{-(\delta-2\delta_c)^2/2\sigma^2} \, d\delta = p_0(> \delta_c). \tag{29}$$

Miraculously, this is just the original PS term: the analysis of Bond *et al.* supplies the PS factor of 2 without needing to cheat. However, since most filter functions that one might want to consider *are* differentiable, this is a rather special case.

The above results show that the solution of the original PS problem depends to an unsatisfactory extent of the filter function used. For sharp k-space filtering (and thus a spatial filter with large-amplitude positive and negative sidelobes), we recover the PS result, with a low-mass slope of

$$M^2 f(M) \propto M^{(n+d)/2d}, \tag{30}$$

whereas Gaussian filtering yields

$$M^2 f(M) \propto M^{\sqrt{(n+d)/2d^2}/\pi}. \tag{31}$$

Thus, in 3D, Gaussian filtering predicts more low-mass objects than PS unless $n < -2.8$. As we have argued above, a generally flatter slope probably also applies for other differentiable filters.

4.4 Density peaks

To solve the connectivity problem in PS, it is necessary to look in more detail at how structures form in a random density field. One powerful model assumes that objects form around *maxima* in the linear density field, and this has generated considerable interest in recent years (Peacock & Heavens 1985; Bardeen *et al.* 1986 [BBKS]; Heavens & Peacock 1988).

Armed with expressions for the number density of peaks, and $M(R_f)$, it is possible to follow the PS route and look at the fraction of mass associated with peaks above the collapse threshold as a function of scale. This yields a mass function with the same low-mass slope as PS, but modified around M_* (Bond 1989). However, as with PS, there is a normalisation problem. In one dimension, the number density of all peaks is $(2\pi R_*)^{-1}$ (see BBKS for definitions of the spectral parameters $_*$ and γ), which encloses a fraction

$$f_{pk} = (2\pi)^{-1/2} \frac{R_f}{R_*} = \left(\frac{3+n}{4\pi} \right)^{1/2} \tag{32}$$

for Gaussian filtering with $M = (2\pi)^{1/2} \rho_0 R_f$, and ignoring peak overlap. In three dimensions, the corresponding expression (for $M = (2\pi)^{3/2} \rho_0 R_f^3$) is

$$f_{pk} = 0.016 R_*^{-3} (2\pi)^{3/2} R_f^3 \simeq \left(\frac{n+5}{15} \right)^{3/2}. \tag{33}$$

These fractions are $\lesssim 0.5$ for power spectra of practical interest. What happens to the mass outside peaks?

We can attack this problem in much the same way as we did for field points. Material which lies outside a peak on one scale can sometimes lie inside a peak on a larger filtering scale. For very low thresholds, the probability of this occurring clearly goes to unity. Thus, as with field points, we can write an expression which is automatically normalised:

$$p(> R_f) = f_{pk}(\delta > \delta_c) + (1 - f_{pk}) p_{up}. \tag{34}$$

The critical term p_{up} is the analogue of the upcrossing probability: the probability that a point outside a peak will find itself inside one on some larger filtering scale. Following our previous reasoning, we again make the ansatz that p_{up} may be estimated by saying that random fields should be effectively independent after some increment Δ in $\ln R_f$. However, there is no reason to suppose that the appropriate value of Δ for this problem will be the same as the one we deduced for the case of field points.

For a small amount of smoothing, almost all peaks will be at essentially the same positions and the number above a threshold will change only by a fractional amount of order $\delta R_f / R_f$. However, *all* these peaks will be more massive, owing to the increase in R_f. Thus, to first order in δR_f, the only change in the fraction of mass which has been inside a peak is due to the mass swept up as the peaks 'bloat'. Since we can find the initial rate of change
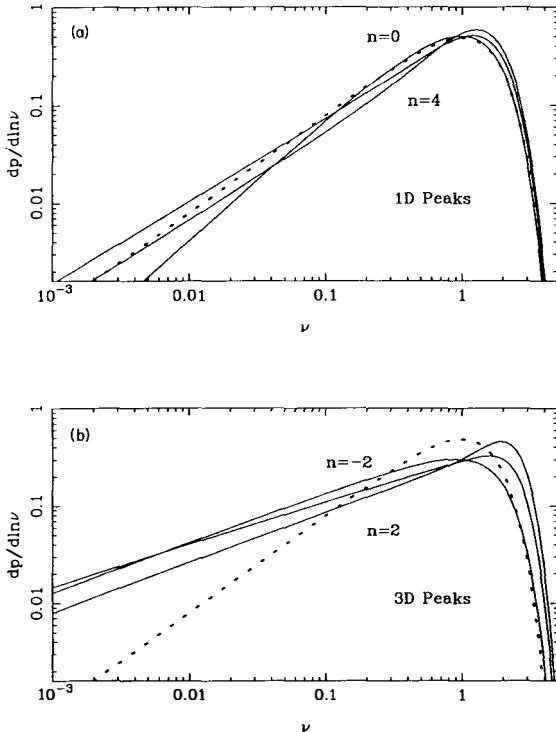
Figure 5.   *The mass functions for peaks in differential form, $dp/d\ln\nu$ (the analogue of Fig. 4 for field points).*
(a) in one dimension
(b) in three dimensions
Again, the PS result is shown dotted. Note that the numbers of low-mass objects predicted are lower than in the field-point case.

of this fraction through the above 'bloating' argument, the appropriate value of Δ may be deduced (Peacock & Heavens 1990):

$$\Delta_{pk} = -\frac{[1 - f_{pk}(0)]\,\ln[1 - f_{pk}(0)]}{d\,f_{pk}(0)}. \tag{35}$$

For the small values of $f_{pk}(0)$ typical in three dimensions, this tends to $\Delta_{pk} \simeq 1/3$, which is roughly an order of magnitude less that Δ_{field}.

We show the results of the above procedure for Gaussian-filtered power-law spectra in Figure 5, in the form of $dp/d\ln\nu$.

Comparison with the field-point results show that far fewer low-mass objects are now predicted. In one dimension, these peak $f(M)$ curves resemble the PS formula, although in three dimensions there are generally still more low-mass objects than predicted by PS. The precise slopes of these curves should not however be taken as having too much significance. The asymptotic behaviour resulting from dp_{up}/dM is

$$M^2 f(M) \propto M^{-\frac{\ln[1-f_{pk}(0)]}{d\Delta}} = M^{f_{pk}(0)/[1-f_{pk}(0)]}, \tag{36}$$

which will change according to the mass estimate we use. For field points, the low-mass slope depended on the analytic form of the filter function. For peaks, this seems less important, but the question of the $M(R_f)$ relation (which is non-trivial for all but spherical top-hat filters) becomes crucial. As we have seen, this formalism generally predicts more low-mass objects than PS. However, this is only true if the above mass dependence is less rapid than the PS value $M^{(n+d)/(2d)}$. Thus, if

$$f_{pk}(0) > \frac{n+d}{n+3d}, \tag{37}$$

then the low-mass part of the mass function will lie *below* that predicted by PS. Equality of slope requires a mass larger than $(2\pi)^{3/2}\rho_0 R_f^3$ by only a moderate factor: 1.7 for $n = 0$.

Is this reasonable? There are a variety of choices for peak mass estimates. The simplest is just the volume of the filter function $M = (2\pi)^{3/2}\rho_0 R_f^3$. Note that the collapse of this volume only requires the filtered $\delta > \delta_c$ at *one* point – in this respect artificial and physical filtering differ markedly. The comparison with the spherical filter provides an alternative method of assigning a mass to peaks. In some sense, the equivalent sphere to (say) a Gaussian filter should be the sphere (of radius R_{sph}) which produces the same rms as the applied filter of scale R_f. In k-space, the Gaussian and spherical filters agree to second order in k if $R_{sph} = \sqrt{5} R_f$. The volume of this sphere is

$$V = \frac{4\pi}{3} 5^{3/2} R_f^3, \tag{38}$$

which is a factor 2.97 larger than $(2\pi)^{3/2} R_f^3$. Although the detailed $R_{sph}(R_f)$ relation for equal rms depends on the power spectrum, it is within 20 per cent of this factor for $n < 0.5$, and so there might be some justification for taking a uniformly larger mass.

An alternative viewpoint is that non-linear development of small-scale perturbations into collapsed objects may be thought of as acting in a sense like a genuine physical filter on the remaining linear portion of the power spectrum. This leads us to consider the problem of the density profile of a peak. Peacock & Heavens (1985) attempted to produce a peak mass estimate by modelling the peaks as triaxial spheroids, and estimating the volume of the region with $\delta > 0$. According to this prescription, there is a distribution of masses for a given ν; however, we shall not be particularly interested in this distribution as any dispersion in mass at a given R_f will be dominated by the fact that we are considering a wide range in

R_f. The mean value of $\log M$ may be approximated (for values $0.5 \lesssim \gamma \lesssim 0.8$ of practical interest) by

$$M = \frac{2^{3/2} \frac{4\pi}{3} \rho_0 R_*^3}{(\gamma^3 + (\frac{0.9}{\nu})^{3/2})}. \tag{39}$$

For a spherical peak with quadratically varying overdensity, the bound mass is a factor $(5/3)^{3/2}$ larger than the above expression.

In short, there may well be grounds for wanting to adopt a peak mass other than the strict $M = (2\pi)^{3/2} R_f^3$ relation, which we have seen can have important consequences for the mass function.

Comparing our mass functions with the simulations of Efstathiou *et al.* (1988b) provides some encouragement for the peak models. For small n, the N-body functions undershoot the peak of the PS distribution by a factor ~ 1.4; the peak models which use the simple mass estimate undershoot by a larger factor than this, but the trend is qualitatively in the right direction. The N-body simulations at present lack the dynamic range to establish the low-mass slope directly, but simple considerations of probability tell us that there must be an excess of low-mass objects if there is a deficit around M_*.

In summary, if we believed that the appropriate mass to assign to a peak is indeed rather larger than the filter volume, the mass function may not be so different from the PS result. The theoretical grounds for such a choice are however hardly compelling; if the PS curve resembles what is seen in N-body simulations, this must look more than ever like a happy accident.

4.5 Conditional multiplicity functions

The trajectory approach allows some interesting extensions. Given an object of mass M_0 at epoch a_0, one can ask what masses existing at some earlier epoch a_1 have merged to produce the final object – this is the conditional multiplicity function. In the language of trajectories, with all variances expressed in terms of linear values at $a = 1$, we are given that a trajectory crosses δ_c/a_0 at mass M_0, and require the distribution of masses M_1 at which the higher threshold δ_c/a_1 is crossed. For the Bond *et al.* random walk model, this is a simple generalisation to two absorbing barriers, and the solution is (Efstathiou 1990):

$$f(M_1, a_1 | M_0, a_0) = \left(\frac{2}{\pi}\right)^{1/2} \frac{\delta_c \sigma_1 (a_0 - a_1)}{a_0 a_1 (\sigma_1^2 - \sigma_0^2)^{3/2}} \left(-\frac{\partial \sigma_1}{\partial M_1}\right) \times$$
$$\exp\left[-\frac{\delta_c^2 (a_0 - a_1)^2}{2 a_0^2 a_1^2 (\sigma_1^2 - \sigma_0^2)}\right]. \tag{40}$$

This equation shows that massive groups at late times are formed from unusually massive groups at early times. If we think of M_0 as representing an Abell cluster at $a_0 = 1$ and go back to $a_1 \simeq 0.25$ when M_* corresponds to a massive galaxy, this shows quantitatively that massive galaxies form early in clusters. In section 6.2 we shall study the 'natural bias' mechanism whereby this tendency, plus a hypothesised star-formation efficiency which depends on epoch, can lead to a bias in the M/L ratios for clusters.

5 Dissipation and galaxy mass functions

The interpretation of mass functions is fairly straightforward when considering only 'haloes' of collisionless dark matter, but is more complex for baryonic matter, where we must ask if the matter has been able to *dissipate* and turn into stars. This question was analyzed in a classic paper by Rees & Ostriker (1977), and has been reconsidered in the context of CDM by Blumenthal *et al.* (1984). Essentially, the point is that in the early universe, all forming structures *can* cool: a set of merging objects will create stars from the gas between them and subsequently will be identified as a single stellar system. Baryonic substructure can thus also be erased – unless star formation in the first generation of the hierarchy is so efficient that all gas immediately becomes locked up in compact $\sim 10^6 \, M_\odot$ clumps of stars. Clearly this did not happen, and one has to consider all manner of complications such as mass loss from stars and the effect of supernova-driven winds on galactic gas (*e.g.* White 1989).

A method for inserting these considerations into the PS prescription was suggested by Peacock & Heavens (1990): one simply assumes that star formation at all epochs has such a low overall efficiency that most baryonic material is always in the form of gas. This is certainly what we see today in clusters of galaxies. Now, in order for dissipation to occur, the redshift of collapse clearly needs to be sufficiently large that there is time for an object to cool between its formation at redshift z_{cool} (when $\delta\rho/\rho \simeq \delta_c$) and the present epoch. More massive objects take longer to cool, and so a simple way of implementing cooling in the PS model is to use a *mass-dependent* threshold: the mass exceeds M if $\delta(M) > \delta_c(1 + z_{cool}(M))$.

The cooling function for a plasma in thermal equilibrium has been calculated by Raymond, Cox & Smith (1976). For an H + He plasma with $Y = 0.25$ and some admixture of metals, their results for the cooling time ($t_{cool} \equiv 3kT/2\Lambda(T)n$) may be approximated as

$$t_{cool}/\text{years} = 1.8 \times 10^{24} \left(\frac{\rho_B}{M_\odot \text{Mpc}^{-3}} \right)^{-1} \left(T_8^{-1/2} + 0.5 f_m T_8^{-3/2} \right)^{-1}, \tag{41}$$

where $T_8 \equiv T/10^8 K$. The $T^{-1/2}$ term represents bremsstrahlung cooling and the $T^{-3/2}$ term approximates the effects of recombination radiation. The parameter f_m governs the metal content: $f_m = 1$ for solar abundances; $f_m \simeq 0.03$ for no metals. Now, the density after collapse is some multiple f_c of the background density at virialisation:

$$\rho_c = f_c \, \rho_0 \, (1 + z_c)^3. \tag{42}$$

Using $\rho_0 = 2.78 \times 10^{11} \Omega h^2 \, M_\odot \text{Mpc}^{-3}$, we obtain

$$t_{cool}/H_0^{-1} = 660 \, (f_c \Omega_B h)^{-1} \, (1 + z_c)^{-3} \, \left(T_8^{-1/2} + 0.5 f_m T_8^{-3/2} \right)^{-1}. \tag{43}$$

The virialised potential energy for constant density is $3GM^2/(5r)$, where the radius satisfies $4\pi\rho_c r^3/3 = M$. This energy must equal $3MkT/(\mu m_p)$, where $\mu = 0.59$ for a plasma with 75% Hydrogen by mass. Hence

$$T/K = 10^{5.1}(M/10^{12} M_\odot)^{2/3} \, (f_c \Omega h^2)^{1/3} \, (1 + z_c). \tag{44}$$

The criterion for formation of a galaxy by the present is that the cosmic time since z_c is some multiple f_t of t_{cool} (for example, a spherical body in quasi-hydrostatic equilibrium has $f_t = 2$; Rees & Ostriker 1977). So, for $\Omega = 1$, we must solve

$$f_t\, t_{cool} = \frac{2}{3} H_0^{-1} (1 - (1 + z_c)^{-3/2}).$$

(45)

If only recombination cooling was important, the solution to this would be

$$(1 + z_c) = (1 + M/M_{cool})^{2/3},$$

(46)

where

$$M_{cool}/M_\odot = 10^{13.1}\, f_t^{-1} f_m f_c^{1/2} \Omega_B \Omega^{-1/2}.$$

(47)

For high metallicity, where bremsstrahlung only dominates at $T \gtrsim 10^8 K$, this equation for z_c will be a reasonable approximation up to $z_c \simeq 10$, at which point Compton cooling will start to operate. Given that we expect at least some enrichment rather early in the progress of the hierarchy, we shall keep things simple by using just the above expression for z_c.

The effects of this cooling threshold on the mass function are illustrated in Fig. 6. At low masses, $z_{cool} \simeq 0$ and the mass function is unaffected by cooling. At high masses, the cooling time is long for masses of order M_c, and $M^2 f(M)$ peaks at a much smaller mass. The sensitivity of this characteristic mass to M_{cool} is not very high: for $n = 0$, it changes by a factor ~ 10 when M_{cool} is altered by a factor 100. Note that including bremsstrahlung makes little difference, since the virial temperatures of the objects near the peak in $M^2 f(M)$ is below 10^8 K. Since the mass functions with and without cooling coincide for low masses, but given that cooling of massive objects is ineffective, probability in the mass function must accumulate at intermediate masses. Thus, the numbers of faint galaxies relative to bright are decreased. If $M_{cool} \ll M_c$, then there is a power-law region between these two masses which differs from the PS slope: $M^2 f(M) \propto M^{\left(\frac{n+3}{6}\right) + \frac{2}{3}}$; i.e. there is an effective change in n to $n + 4$.

We should not claim too much from the above analysis, as several potentially important points are neglected. First, erasure of baryonic sub-structure may be imperfect, which would lead us to underestimate the numbers of low-mass objects (White & Rees 1978). Second, the criterion of equating cooling time with look-back time will work only if an object is able to cool undisturbed over this time; if subsequent generations of the hierarchy collapse while the object is still cooling quasi-statically, then the gas will be reheated and collapse may never occur (see White & Rees for this point also). Objects are immune to this effect only if the cooling time is shorter than the free-fall time, which turns out to be simply a criterion on mass of the same order as M_{cool} (see e.g. Efstathiou & Silk 1983). Nevertheless, the qualitative point that cooling should lead to a steepening of the luminosity function, simply through probability conservation, is an important one, and may well help to reconcile the observed galaxy luminosity function with the PS low-mass slope.
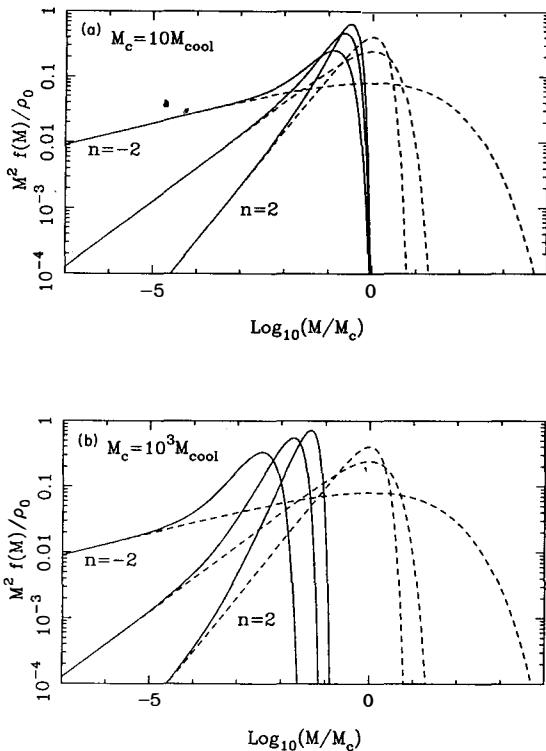
Figure 6. *The mass functions resulting from the PS formalism with the adoption of the cooling threshold criterion ($\delta_c \rightarrow \delta_c(1 + M/M_{cool})^{2/3}$). The solid lines show the mass functions with cooling, and the dashed lines show the mass functions without cooling; the two sets of curves coincide only at very low masses. M_c is the mass scale with $\sigma(M_c) = \delta_c$.*
    Note (a) the relative insensitivity to M_{cool} and (b) that over several orders of magnitude in mass, the relative numbers of low-mass objects are greatly reduced by cooling.

6 Biased mass functions

We now need to consider the difficult step from mass functions to luminosity functions. This is complicated by the now common belief that galaxy formation is biased: mass does not trace light.

Why do we believe this? Cluster blue M/L ratios of about $300h$ (*e.g.* Merritt 1987) or the field galaxy velocity dispersion of about 300 kms^{-1} (Bean *et al.* 1983) yield an apparent $\Omega \simeq 0.1 - 0.2$ if mass traces light. Any theoretical belief in $\Omega = 1$ 'proves' the necessity for an extra uniform mass component. Some observational backing for the idea of a high-density

universe has been forthcoming recently through the comparison between redshift-survey and peculiar-velocity data (*e.g.* Yahil 1989) – see also the contributions by Bertschinger and Yahil at this meeting. Even in the absence of evidence for $\Omega = 1$, however, there are some arguments which require bias, concerning whether one can produce the observed clustering pattern purely via gravitational instability. One of the early arguments for bias was the difficulty in producing the observed contrast in density between voids and superclusters in N-body simulations of CDM universes (White *et al.* 1987a). A more general investigation has been performed by Weinberg (1989), who considers redshift-survey data and attempts to reconstruct linear initial conditions from which the observed configuration could have arisen. It is difficult to achieve this in an unbiased model, even for low values of Ω. The problem is always the same: to empty the voids sufficiently, the models must be evolved to a point where the superclusters break up into nonlinear clumps. This conclusion arises independently of information on velocity perturbations.

6.1 The threshold model

The dominant mechanism for achieving bias has been what is known as the threshold model (Davis *et al.* 1985; BBKS). This model appears to have caused some confusion, as it is inspired by a phenomenon which has nothing whatsoever to do with bias – the excess clustering of Abell clusters. Although more recent determinations have weakened original estimates of the strength of the effect (Sutherland 1988), it still appears that $\xi(r)$ for richness 1 clusters is about an order of magnitude larger than $\xi(r)$ for galaxies. The explanation for this is almost certainly the one suggested by Kaiser (1984): Abell clusters are the most massive virialised structures in existence today, and so they must correspond to high peaks in the mass distribution when filtered on scales ~ 1 Mpc.

For very high peaks, this density threshold amplifies any large-scale density perturbations, leading asymptotically to

$$\xi_{pk}(r) \simeq \nu^2 \frac{\xi(r)}{\xi(0)} \tag{48}$$

(Kaiser 1984; BBKS). So, clusters do not describe the large-scale mass distribution, but this is not what we mean by bias – the effect still operates even if mass traces light. Nevertheless, this shows a way in which galaxy clustering could also fail to trace mass, if galaxies were to form only at high peaks of the density field smoothed on ~ 100 kpc scales. The two thresholds have a very different status, however. For clusters, the threshold arises because low-ν fluctuations have not yet had time to collapse; when they do, the 'bias' in clustering in erased (Cole & Kaiser 1989). For galaxies, some way has to be found of making the early-forming objects cause the later ones to be 'still-born'. No convincing way of propagating disruptive signals over the required distances has been found, however (*e.g.* Dekel & Rees 1987).

Apart from physical difficulties with its implementation, the threshold model is also rather inconsistent with the PS philosophy. So far, we have been assuming that low-ν fluctuations would make objects of low mass, whereas the threshold model considers that they

will make 'failed' galaxies. This is certainly a radical consequence of the threshold model, and these dark potential wells have been appealed to *e.g.* in cases of gravitational lensing where an obvious galaxy lens is lacking (Subrahmanian, Rees & Chitre 1987). However, it seems more likely that this conclusion arises from the unrealistic limitation of trying to characterise all galaxies by a single filter radius. Similarly, extensions of the threshold model to produce 'morphological' luminosity functions where different Hubble types are associated with different ν values (*e.g.* Martínez-González & Sanz 1988; Evrard 1989a) are probably to be viewed with some scepticism. Initial motivation for this viewpoint was the hope that high peaks would be spun up less by tidal torques and would then form ellipticals (Blumenthal *et al.* 1984); however, this turned out not to be the case in practice (Heavens & Peacock 1988).

Extensions of the threshold model have been attempted (*e.g.* Borgani & Bonometto 1990), in which the threshold is made fuzzy. However, this only makes a significant difference to things if the width of the 'selection function' is so large that some objects form with $\delta \ll 1$. In the context of the PS approach, it is very hard to see how this can come about.

6.2 Natural bias

A more promising candidate is so-called 'natural' bias (White *et al.* 1987b; Kaiser 1988; Cole & Kaiser 1989). Here, one hypothesizes that star-formation efficiency may be a function of epoch, so that a halo of mass M collapsing at redshift z produces a stellar luminosity

$$L \propto M^\alpha (1+z)^\beta. \tag{49}$$

The model parameters $\alpha = 4/3$, $\beta = 2$ produce a perfect theoretical 'Faber-Jackson' relation $L \propto \sigma_v^4$, hence the name 'natural'.

The important point to note is that, in this model, there are no 'failed' galaxies: clusters have dark haloes. In this respect, natural bias ends up looking much the same as pictures in which the bias arises through dynamical friction (West & Richstone 1988; Carlberg, Couchman & Thomas 1990). Thus, when we come to convert from mass functions to luminosity functions, observed cluster M/L ratios are *not relevant*. PS-type mass functions are normalised to account for the total density of the universe; similarly, $\phi(L)$ accounts for the total luminosity density. There is therefore no choice but to adopt the ratio

$$M/L \simeq 2000 \Omega h \tag{50}$$

(in blue solar units; Efstathiou *et al.* 1988a). This predicts a surprisingly large mass for M_* galaxies: $M_* = 10^{13.2} \Omega h^{-1} \, M_\odot$, but this should not be seen as a problem. PS allocate all the mass to one object or other, but most of it lies in the outer parts of haloes, where it is so loosely bound as to be unobservable. Studies such as Bond (1989) and Colafrancesco *et al.* (1989) which use observed M/L values are therefore inconsistent: integration under their luminosity functions will not yield the observed luminosity density.
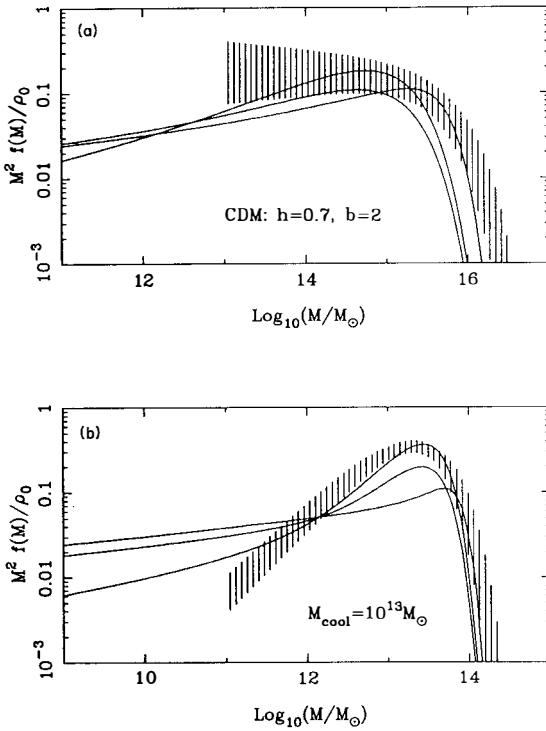
Figure 7. *The predicted mass functions for CDM with closure M/L, using Gaussian filtering with $M = (2\pi)^{3/2}\rho_0 R_f^3$ and $\delta_c = 1.69$. (a) shows the mass function for 'haloes'. Working from the right at the high-mass end, the three solid curves show the peak ansatz, PS, and the ansatz for field points. The hatched area is the group multiplicity function from Bahcall (1979). (b) is the same as (a), but for 'cooled' objects. The hatched area is the fit to the galaxy luminosity function from Efstathiou et al. (1988a).*

In fact, mass dependence of M/L has little impact on predicted luminosity functions. The PS prescription under natural bias would predict a low-mass slope

$$f(L) \propto L^{-2 + \frac{n-d}{2d\alpha + \beta(n+d)}}, \tag{51}$$

which will cause a change in slope of only $\lesssim 0.2$ in practice. We therefore show the mass functions predicted for CDM under the assumption of constant M/L in Figure 7.

7 Applications

The important applications of an understanding of cosmological mass functions come both at

the present epoch and in looking at conditions predicted for high redshifts. Data on evolution of luminosity functions for clusters and galaxies are now becoming available that can yield powerful constraints on models for structure formation.

7.1 Cluster mass functions

The abundance of rich clusters of galaxies is a fundamental test of the gravitational instability picture. Given that the effective spectral index on $\sim$ Mpc scales is fixed by observation as $n_{\text{eff}} \simeq -1$, the abundance of clusters is determined once the amplitude of fluctuations is set via a normalisation on large scales. A common scheme is to specify that the rms density fluctuation in spheres of radius $8h^{-1}$ Mpc is $1/b$, where b is the linear bias parameter. Since b can be constrained directly from peculiar-velocity data, there is very little freedom of maneuver in the allowed abundance of rich clusters. If there were too many of these, we would have to abandon a fundamental assumption and move to non-Gaussian fluctuations, or even reject an origin in gravitational instability. The observational situation is surveyed by Frenk *et al.* (1990), who conclude that there are indeed apparently too many clusters of very high velocity dispersion to be consistent. However, by analysis of simulated cluster catalogues selected from a CDM simulation, they show that this excess plausibly consists of spurious alignments of objects of lower richness. The conclusion of Frenk *et al.* is that the data are consistent with a value $b \simeq 2.5$, which they had previously preferred for matching the large-scale clustering pattern. It remains to be seem whether there is an inconsistency here between the values $b \lesssim 1.5$ preferred by direct analyses of IRAS galaxy samples (Kaiser & Lahav 1989; see also the contributions by Bertschinger and Yahil at this meeting).

The constraints from clusters should become stronger at high redshift; M_* shifts to lower masses and very rich clusters should eventually become very rare. Evrard (1989b) and Peebles, Daly & Juskiewicz (1989) have pointed out that there may be some evidence for an excess of high-σ_v clusters at high redshift. The problem here is that, in arguing on the basis of a few extreme examples, the possibility of the selection biases identified by Frenk *et al.* becomes an especial worry. A better way of proceeding may be to use an object such as a radio galaxy to serve as a marker of a possible distant cluster, and then to count nearby galaxies to estimate the richness. Yates, Miller & Peacock (1989) have shown in this way that the average environment of radio galaxies at $z \simeq 0.5$ is Abell $R \simeq 0$ – certainly no lower than at $z = 0$. However, searches in narrow-band filters for Lyα companions to radio galaxies at $z \gtrsim 1.6$ have been uniformly unsuccessful (Spinrad 1989). Although this results remains to be quantified in terms of cluster richness, this may be the first hint of the expected dearth of distant rich clusters.

7.2 Galaxy mass functions

A more complex problem is concerned with how the galaxy luminosity function may change with redshift. It is reasonably well established from raw number counts that evolution of

CDM h=0.7 b=2 $M_{cool}=10^{13}M_\odot$ z=0-10

$M^2 f(M)/\rho_0$

$Log_{10}(M/M_\odot)$
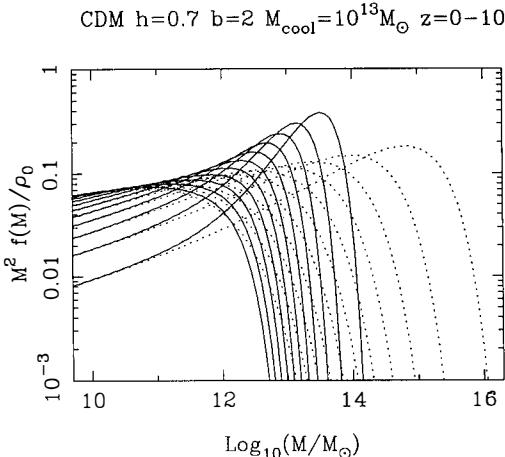
Figure 8. *The epoch dependence of the cooled mass function in CDM from Figure 7b – using the PS version only (dotted lines show the mass function without cooling). Note that, for $z \gtrsim 3$, the characteristic mass becomes $\lesssim M_*$ for galaxies.*

some form is needed, but until recently all attempts to fit this behaviour assumed luminosity functions of unchanging form, with L^* being brighter in the past owing to stellar evolution. Indeed, this assumption of pure luminosity evolution was the basis of the paper by Loh & Spillar (1986), which claimed to estimate $\Omega \simeq 1$ by looking at $\phi^*(z)$.

These assumptions have, however, been overturned by redshift data. Starting with the paper by Broadhurst *et al.* (1988), and continuing with unpublished deeper surveys by the Durham group, it has been established that the luminosity function must change shape with redshift – in the sense that there were relatively more low-luminosity galaxies in the past. The blue population which dominate the excess at faint magnitudes turn out to be not luminous galaxies at $z \simeq 1$, but weaker objects at $z \lesssim 0.5$. Although this is such a new area that many interpretations are possible, these observations do find a natural explanation in hierarchical models for structure formation. Dark haloes continue to merge at all redshifts, so is it reasonable to expect mergers of luminous objects to be important at $z \lesssim 1$? To attempt to get some feeling for this, figure 8 shows how the galaxy mass function varies with epoch in CDM, according to the 'PS+cooling' model of Peacock & Heavens (1990).

On the basis of this simple model, a significant amount of galaxy merging between $z = 1$ and the present appears to be feasible. Note that the curvature of the CDM power spectrum means that indeed the relative numbers of low-mass galaxies is increased at high redshift. However, the change appears to be too small to explain the galaxy redshift-survey data; a more realistic model of merging is probably needed. This is certainly a problem

where much theoretical effort needs to be invested to match the excitement of observational developments.

Finally, some more speculative remarks concerning quasars. It now appears to be becoming established that these are rarer at $z \simeq 4$ than at $z \simeq 2$ – the so-called 'redshift cutoff' (*e.g.* Peacock 1985; Warren *et al.* 1988; Schmidt *et al.* 1988). The theoretical interpretation of this effect is plausibly that outlined by Efstathiou & Rees (1988): at high redshift one expects that the massive galaxies needed to host luminous quasars have not yet been assembled (*cf.* figure 8). Does this mean that we can never hope to see quasars at $z = 10$? Things may not be so hopeless, as a glance at figure 8 will confirm. Beyond $z \simeq 3$, the abundance of objects with $M \gtrsim 10^{13} M_\bullet$ is indeed dropping exponentially; however, the characteristic 'galaxy' mass at $z = 10$ is only some $10 - 30$ times smaller than it was at $z = 3$. By mass conservation, the number of objects of this typical mass has *risen* by the same factor. The number of potential sites for an AGN at high redshift is therefore very large – and one does not need to go impossibly faint to stand a chance of detecting them if quasar luminosity scales as galaxy mass. This begs many questions about black hole formation, of course, but it would still seem worth having a small bet on the $z = 10$ barrier being broken one day.

8 Concluding remarks

I hope I have said enough to indicate the wealth of new developments going on in this subject. The pace of change is such that we can look forward over the next few years to a very fruitful interaction between data and theory.

I thank my collaborators: Alan Heavens, Sergei Shandarin and Brian Williams, joint work with whom is presented above. They are not to be held to blame for some of the more speculative parts of this article.

References

Bahcall, N.A., 1979. *Astrophys. J.*, **232**, 689.

Balian, R. & Schaeffer, R., 1989. *Astr. Astrophys.*, **220**, 1.

Bardeen, J.M., Bond, J.R., Kaiser, N. & Szalay, A.S., 1986. *Astrophys. J.*, **304**, 15. (BBKS)

Bean, A.J., Efstathiou, G., Ellis, R.S., Peterson, B.A. & Shanks, T., 1983. *Mon. Not. R. astr. Soc.*, **205**, 605.

Blumenthal, G.R., Faber, S.M., Primack, J.R. & Rees, M.J., 1984. *Nature*, **311**, 517.

Bond, J.R., 1989. in *Large-Scale Motions in the Universe*, proc. Vatican Study Week.

Bond, J.R., Cole, S., Efstathiou, G. & Kaiser, N., 1989. In preparation.

Borgani, S. & Bonometto, S.A., 1990. *Astrophys. J.*, **348**, 398.

Broadhurst, T.J., Ellis, R.S. & Shanks, T., 1988. *Mon. Not. R. astr. Soc.*, **235**, 827.

Brandenberger, R.H., 1990. in *Physics of the Early Universe*, proc 36th Scottish Universities Summer School in Physics, eds Peacock, J.A., Heavens, A.F. & Davies, A.T. (Adam Hilger), p281.

Burgers, J.M., 1940, *Proc. Roy. Neth. Acad. Sci.*, **43** ,2.

Burgers, J.M., 1974, *The Nonlinear Diffusion Equation.*, Reidel, Dordrecht, H●lland.

Carlberg, R.G. & Couchman, H.M.P., 1989. *Astrophys. J.*, **340**, 47.

Carlberg, R.G., Couchman, H.M.P. & Thomas, P.A., 1990. *Astrophys. J.*, **352**, L29.

Chandrasekhar, S., 1943. *Rev. Mod. Phys.*, **15**, 1. Reprinted in *Selected papers on Noise and Stochastic Processes*, p. 3, ed. Wax, N., Dover.

Colafrancesco, S., Lucchin, F. & Matarrese, S., 1989. *Astrophys. J.*, **345**, 3.

Cole, S. & Kaiser, N., 1989. *Mon. Not. R. astr. Soc.*, **237**, 1127.

Davis, M., Efstathiou, G., Frenk, C.S. & White, S.D.M., 1985. *Astrophys. J.*, **292**, 371.

Dekel, A. & Rees, M.J., 1987. *Nature*, **326**, 455. Terlevich, R.J. & Wegner, G., 1987. *Astrophys. J.*, **313**, 42.

Efstathiou, G., 1990. in *Physics of the Early Universe*, proc 36th Scottish Universities Summer School in Physics, eds Peacock, J.A., Heavens, A.F. & Davies, A.T. (Adam Hilger), p361.

Efstathiou, G., Ellis, R.S. & Peterson, B.A., 1988a. *Mon. Not. R. astr. Soc.*, **232**, 431.

Efstathiou, G., Frenk, C.S., White, S.D.M. & Davis, M., 1988b. *Mon. Not. R. astr. Soc.*, **235**, 715.

Efstathiou, G. & Rees, M.J., 1988. *Mon. Not. R. astr. Soc.*, **230**, 5P.

Efstathiou, G. & Silk, J., 1983. *Fund. Cosm. Phys.*, **9**, 1.

Evrard, A.E., 1989a. *Astrophys. J.*, **341**, 26.

Evrard, A.E., 1989b. *Astrophys. J.*, **341**, L71.

Frenk, C.S., White, S.D.M., Efstathiou, G. & Davis, M., 1990. *Astrophys. J.*, **351**, 10.

Gurbatov, S.N., Saichev, A.I., & Shandarin, S.F., 1985, *Sov. Phys. Dokl.* **30**, 921.

Gurbatov, S.N., Saichev, A.I., & Shandarin, S.F., 1989, *Mon. Not. R. astr. Soc.*, **236**, 385.

Heavens, A.F. & Peacock, J.A., 1988, *Mon. Not. R. astr. Soc.*, **232**, 339.

Henriksen, R.N. & Lachièze-Rey, M., 1990. *Mon. Not. R. astr. Soc.*, in press.

Kaiser, N., 1984. *Astrophys. J.*, **284**, L9.

Kaiser, N., 1988. *Large-Scale Structures of the Universe*, proc. IAU symp. no. 130, eds J. Audouze, M.-C. Pelletan & A. Szalay (D. Reidel), p43.

Kaiser, N. & Lahav, O., 1989. *Mon. Not. R. astr. Soc.*, **237**, 129.

Katz, N & Hernquist, L., 1989. in *The Epoch of Galaxy Formation*, eds Frenk, C.S., Ellis, R.S., Shanks, T., Heavens, A.F. & Peacock, J.A., NATO ASI C264, 433.

Kida, S., 1979 *J. Fluid Mech.*, **93**, 337

Kotov, É.,V., & Shandarin, S.F., 1989, *Soviet Astron.*, **32**, 351

Loh, E.D. & Spillar, E.J., 1986. *Astrophys. J.*, **307**, L1.

Lucchin, F. & Matarrese, S., 1988. *Astrophys. J.*, **330**, 535.

Martínez-González, E. & Sanz, J.L., 1988. *Astrophys. J.*, **332**, 89.

Merritt, D., 1987. *Astrophys. J.*, **313**, 121.

Peacock, J.A., 1985. *Mon. Not. R. astr. Soc.*, **217**, 601.

Peacock, J.A. & Heavens, A.F., 1985. *Mon. Not. R. astr. Soc.*, **217**, 805.

Peacock, J.A. & Heavens, A.F., 1990, *Mon. Not. R. astr. Soc.*, **243**, 133

Peebles, P.J.E., 1980, *The Large Scale Structure of the Universe*, Princeton University Press.

Peebles, P.J., Daly, R.A. & Juskiewicz, R., 1989. *Astrophys. J.*, **347**, 563.

Press, W.H. & Schechter, P., 1974. *Astrophys. J.*, **187**, 425. (PS)

Raymond, J.C., Cox, D.P. & Smith, B.W., 1976. *Astrophys. J.*, **204**, 290.

Rees, M.J. & Ostriker, J.P., 1977. *Mon. Not. R. astr. Soc.*, **179**, 541.

Saslaw, W.C. & Hamilton, A.J.S., 1984. *Astrophys. J.*, **276**, 13.

Schaeffer, R., 1987. *Astr. Astrophys.*, **180**, L5.

Schmidt, M., Schneider, D.P. & Gunn, J.E., 1988. in *Proceedings of a workshop on optical surveys for quasars*, eds Osmer, P.S., Porter, A.C., Green, R.F. & Foltz, C.B, A.S.P. conf. ser., **2**, 87.

Shandarin, S.F., 1987, in *Proc. IAU Symp.* No **104**

Shandarin, S.F., & Zeldovich, Ya.B., 1989, *Rev. Mod. Phys.*, **61**, 185.

Spinrad, H., 1989. in *The Epoch of Galaxy Formation*, eds Frenk, C.S., Ellis, R.S., Shanks, T., Heavens, A.F. & Peacock, J.A., NATO ASI C264, 39.

Subrahmanian, K., Rees, M.J. & Chitre, S.M., 1987. *Mon. Not. R. astr. Soc.*, **224**, 283.

Sutherland, W., 1988. *Mon. Not. R. astr. Soc.*, **234**, 159.

Warren, S.J., Hewett, P.C. & Osmer, P.S., 1988. in *Proceedings of a workshop on optical surveys for quasars*, eds Osmer, P.S., Porter, A.C., Green, R.F. & Foltz, C.B, A.S.P. conf. ser., **2**, 96.

Weinberg, D., 1989. PhD thesis, Univ. of Princeton.

Weinberg, D. & Gunn, J.E., 1990. *Astrophys. J.*, **352**, L25.

West, M.J. & Richstone, D.O., 1988. *Astrophys. J.*, **335**, 532.

White, S.D.M., 1989. in *The Epoch of Galaxy Formation*, eds Frenk, C.S., Ellis, R.S., Shanks, T., Heavens, A.F. & Peacock, J.A., NATO ASI C264, 15.

White, S.D.M. & Rees, M., 1978. *Mon. Not. R. astr. Soc.*, **183**, 341.

White, S.D.M., Frenk, C.S., Davis, M. & Efstathiou, G., 1987a. *Astrophys. J.*, **313**, 505.

White, S.D.M., Davis, M., Efstathiou, G. & Frenk, C.S., 1987b. *Nature*, **330**, 451.

Williams, B.G., Heavens, A.F., Peacock, J.A. & Shandarin, S.F. 1990. *Mon. Not. R. astr. Soc.*, to be submitted. (WHPS)

Yahil, A., 1989. in *Large-Scale Motions in the Universe*, proc. Vatican Study Week.

Yates, M.G., Miller, L. & Peacock, J.A., 1989. *Mon. Not. R. astr. Soc.*, **240**, 129.

Zeldovich, Ya. B., 1970, *Astron. Astrophys.*, **5**, 84.

COSMIC STRUCTURES: FROM MASS TO LIGHT

Alain BLANCHARD
DAEC, Observatoire de Meudon, 92195 Meudon, France.

David VALLS–GABAUD
Institut d'Astrophysique de Paris, 98bis Bld Arago, 75014 Paris, France
Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, England

Gary MAMON
DAEC, Observatoire de Meudon, 92195 Meudon, France.

Summary: We investigate possible approaches to the mass function of cosmic structures and find that a general expression can be established, enabling us to understand the validity of the Press and Schechter formalism (1974). If fragmentation of non-linear objects is negligible and the initial fluctuations are scale-invariant, then the value of the low-mass slope given by the PS expression is an upper limit. Taking into account the effect of cooling, we find that cooling was too efficient to produce the present number of galaxies if no heating mechanism is present. If heating is related to the galaxy formation rate, then the luminosity function has a slope close to -1, in agreement with observations.

1 The mass function of cosmic structures

Analytical derivations of the mass function of cosmic structures meet several difficulties (*e.g.*, Press and Schechter, 1974; Peacock and Heavens, 1990). For instance, it is necessary to know how a non-linear structure can be embedded into a larger one. A second problem is that an object that becomes non-linear can fragment into smaller structures. In this case, the mass function can be entirely dominated by fragmentation processes (*e.g.*, Newman and Wasserman, 1990) so we bypass this uncertainty by assuming that fragmentation is negliglible. Let $\mathcal{F}_{NL}(> m)$ be the fraction of mass that is in non-linear structures with masses larger than m. The mass function $N(m)$ is related to $\mathcal{F}_{NL}(> m)$:

$$N(m)\ dm\ =\ -\ \frac{\bar{\rho}}{m}\ \frac{d\mathcal{F}_{NL}(>m)}{dm}\ dm \tag{1}$$

where $\bar{\rho}$ is the present mass density of the universe. It is possible to write a general form for $\mathcal{F}_{NL}(> m)$ in terms of the initial fluctuation spectrum. Consider an infinite set of random spheres that overlay all the masses in the Universe contained in the initial fluctuations. Let

$$\int_{\nu_S}^{+\infty} F_R(\nu)\, d\nu$$

be the fraction of the mass that is embedded in spheres of radius R, with mean density contrast larger than $\nu_S\sigma$, where σ is the r.m.s. linear density contrast inside a sphere of radius R. This quantity is different from the probability that the spheres have a density above $\nu_S\sigma$. Without loss of generality, we can introduce a selection function $s_R(\delta)$ that gives the fraction of mass in spheres between ν and $\nu + d\nu$ that has actually virialised by now. In the absence of fragmentation, this mass will turn into objects of masses larger than $m(R)$. $\mathcal{F}_{NL}(> m)$ is therefore

$$\mathcal{F}_{NL}(>m)\ =\ \int_{-\infty}^{+\infty} s_R(\nu\sigma)\, F_R(\nu)\, d\nu \tag{2}$$

This equation is almost too simple, but it has the virtue of being *exact*, and of solving the *cloud-in-cloud* problem. The unknown behaviour of non-linear development is described by the function s_R . The mass m associated with spheres of radius R can be written

$$m\ =\ \bar{\rho}\, A(\nu)\, R^3$$

This comes as a consequence of the assumption that fragmentation is negligible.

The standard gravitational instability scenario (assuming a Friedman-Lemaître model with $\Omega = 1$) provides the asymptotic behaviour of the selection function. The transition

from the linear to non-linear regime is controlled by the density fluctuation $\delta\rho/\rho$ of the perturbation:

$$(\text{Non-linear}) \iff \delta\rho/\rho \geq \delta_s$$

As high density contrast fluctuations necessarily produce non-linear objects, the selection function must tend towards $1(0)$ when δ tends towards plus (minus) infinity. This implies that the derivative $s'_m(\delta)$ is null outside a finite domain. The key assumption here is that the function s_R goes from zero to one on scales that typically satisfy the criteria $\delta \approx \delta_S$, *i.e.*, the transition from the linear to non-linear regime is controlled *only* by the r.m.s. amplitude of the fluctuation.

The dependence on smoothing scale (Eq. 2) will never enable us to find a general derivation of the mass function. However, for scale-invariant fluctuations there are several simplifications. First, the distribution function F no longer depends on the smoothing length R. Second, the selection function does not depend on R, and third, the mass m scales as R^3 in the limit $\nu \to 0$. The mass function can then be derived as:

$$N(m)\ dm\ =\ -\frac{\bar{\rho}}{m}\frac{1}{\sigma^2}\frac{d\sigma}{dm}\int_{-\infty}^{+\infty} s'(u)\,F(\nu)\,u\,du\ dm \tag{3}$$

2 The low-mass regime in hierarchical models

The behaviour of the mass function in the low-mass regime can be derived from Eq. (8). As we have mentioned above, s' is zero outside a finite domain. Therefore, in the small-mass limit, the mass function is:

$$N(m)\ dm\ =\ -\frac{\bar{\rho}}{m^2}\frac{1}{\sigma}\frac{dln(\sigma)}{dln(m)}\,F(0)\int_{-\infty}^{+\infty} s'(u)\,u\,du\ dm \tag{4}$$

Therefore, the mass function *on these scales* is only fixed (up to a constant factor) by the variation of the r.m.s fluctuation with mass, and therefore the *low-mass dependence of the mass function is correctly given by the PS behaviour*.

It follows that a consistent derivation of the mass function is possible and that the standard criterion for non-linearity, based on a threshold of the linear density contrast, automatically leads to a mass function with the same low-mass slope as the PS mass function, assuming scale-invariant fluctuations and the absence of fragmentation. It is then tempting to conclude that the mass function should in fact have this behaviour. However, Peacock and Heavens (1990) have pointed out one of the oversimplifications of the PS approach. In this latter description, a fluctuation on a given linear mass scale is assumed not to contribute to a virialised object of larger mass. This may be incorrect because a linear fluctuation can be embedded in a larger non-linear one. We expect such an effect to be small, since linear fluctuations should usually lie outside non-linear regions rather than inside, but this effect can indeed change the low-mass regime. However,

$$\int_{\nu_s}^{+\infty} F(\nu)\,d\nu$$

is certainly the minimal fraction of mass that is embedded in objects with mass greater than m, therefore *it is an underestimation of* $\mathcal{F}_{NL}(>m)$ *and in consequence the fraction of mass in smaller objects should be lower than the PS one*.

$$\Omega_b = 0.1 \qquad h = 0.5 \qquad \delta_c = 1.68 \qquad f_H = 0.1$$
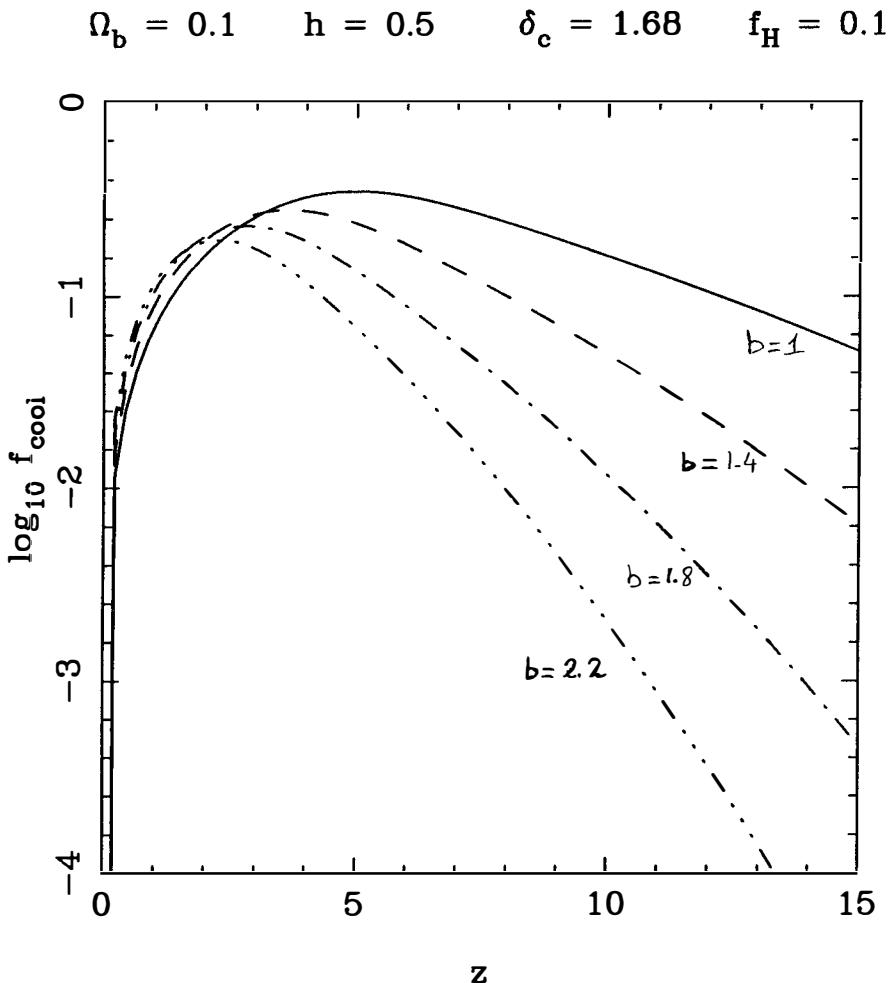

Figure 1: The fraction of potentially cooled gas against redshift. The different curves correspond to different normalisations of the CDM sprectrum. This represents the *minimal* fraction of the gas that should have been processed in stars by now.

3   Gas evolution

Clearly, in order to interpret the meaning of the galaxy luminosity function it is necessary to understand under which conditions the primeval gas has been processed into stars, and its implications for the luminosity of haloes. A simple analysis of star formation processes is sufficiently helpful. In the top-hat model for spherical collapse, the virial temperature can be derived, assuming that the structure virialises at half the radius at maximum expansion, as

$$\left(\frac{T_v}{10^4 K}\right) \quad = \quad 15.5 \ (1 + z) \ \left(\frac{M}{10^{12} M_\odot}\right)^{2/3} \tag{5}$$

At each epoch, only objects with masses in the range $M_1(z)$ to $M_2(z)$ are able to cool in less than one Hubble time. Therefore, the fraction $f_{cool}(z)$ of gas in cosmic structures that is able to cool in less than a Hubble time is

$$f_{cool}(z) \quad = \quad \frac{1}{\rho} \int_{M_1(z)}^{M_2(z)} N(m) \, m \, dm$$

Figure 1 presents f_{cool} computed with the CDM spectrum, using the PS formalism as modified by Schaeffer and Silk (1985). The bias parameter b_ρ is defined as

$$\frac{1}{b_\rho} \quad = \quad \sigma(8h^{-1} \text{Mpc})$$

As expected, cooling is more efficient at high reshift for low b_ρ. Figure 1 also shows the evolution of f_{cool} when the cooling time is one tenth of the Hubble time, but this makes only a small difference since cooling is very efficient and usually $t_{cool} \ll t_H$ at any redshift. The cooled gas must remain in galaxies as the fraction of diffuse neutral gas is known to be very small from the Gunn-Peterson test. We therefore expect that the maximum value of $f_{cool}(z)$ approximates the minimal fraction of the gas that has hitherto cooled and been processed in galaxies. We can now compare f_{cool}^{max} with observational limits. First, f_{cool}^{max} represents the minimal value of the fraction of the mass associated with galaxies. The minimum value is justified by the fact that galaxies can contain baryons whose outer parts have not yet cooled. Therefore

$$\Omega_g \quad \geq \quad f_{cool}^{max}$$

where Ω_g is the contribution from galaxies to Ω_0 . This leads to

$$f_{cool}^{max} \quad \leq \quad 0.01 \, \text{to} \, 0.03$$

Secondly, the contribution of stars to the total density is

$$\Omega_* = f_{cool}^{max} \, \Omega_b \quad = \quad \Omega_L \left(\frac{M}{L}\right)_*$$

and therefore

$$f_{cool}^{max} \quad = \quad \frac{\Omega_L \, (M/L)_*}{\Omega_b} \approx .01 - .03 \left(\frac{M}{L}\right)_*$$

$$\Omega_b = 0.1 \qquad h = 0.5 \qquad \delta_c = 1.68 \qquad f_H = 0.1$$
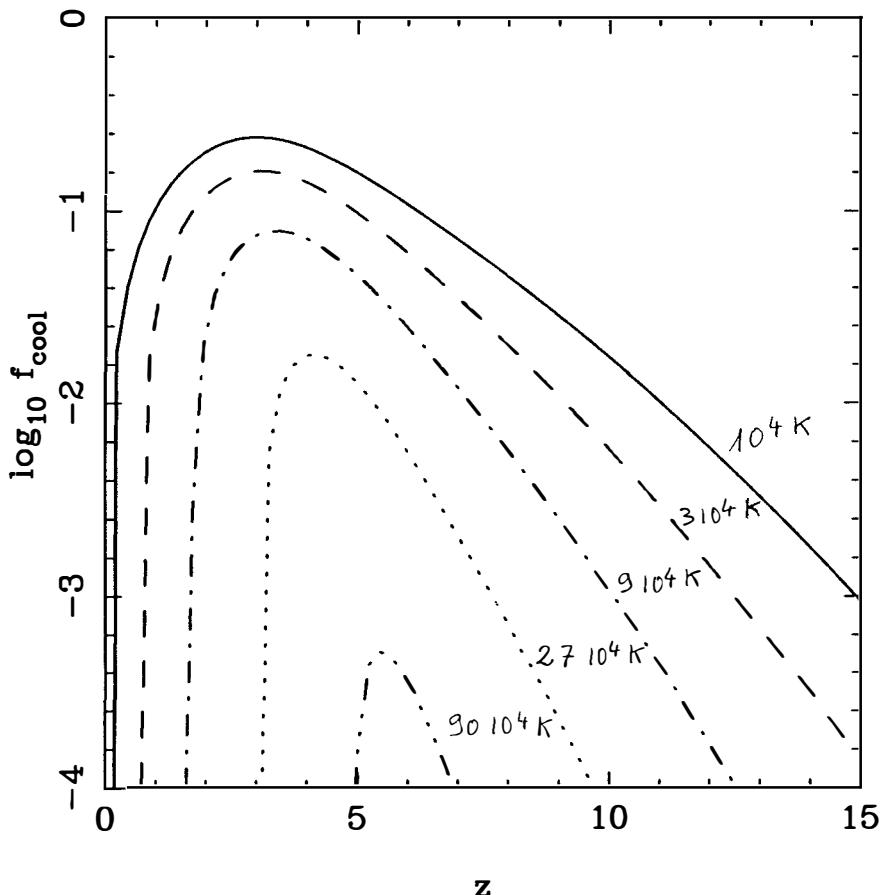

Figure 2: The same quantity than in figure 1, but when the gas is assumed to be maintained at some temperature T_*. The diferents curves correspond to different temperature.

These values are much smaller than the computed ones and we conclude that in a simple CDM model the galaxy formation rate expected when using simple cooling arguments produces far too many galaxies. This is a rather robust conclusion, as the fraction of cooled gas is probably underestimated.

There are two possible solutions to this problem. First, the (M/L) ratios may be underestimated. Dynamical studies of spiral galaxies indicate, however, that $(M/L)_{disk}$ agrees with the M/L ratio of the stellar content, so that the evidence for dark matter *intrinsic* to the disk seems very weak. Unless the production of low-mass stars is concentrated in the outer parts of bright galxies, the fraction of primordial gas that has cooled by now is certainly smaller than 10%.

A more likely solution is that the gas is kept hot by some heating mechanism. This is certainly the simplest way to explain the limits on the fraction of neutral gas implied by the Gunn-Peterson test. The mechanism of heating is not yet identified, but it is likely to be the consequence of the formation of primordial stars, *i.e.*, of cooling, and therefore heating will cease when cooling becomes ineffective. Within this framework, we can determine the temperature at which the gas is heated, in order to produce the observed number of galaxies. In fact, if the gas is heated at a temperature greater than 10^6 K, the cooled fraction becomes negligible (see Fig. 2). Therefore, if there is some efficient heating mechanism, the gas will be heated at a temperature T_* close to this value (Dekel and Silk, 1986). In order to obtain $f \approx 0.01$, a critical temperature of T_* $\approx 3 \times 10^5$ K is needed. As long as the temperature is in this range, haloes that form with a virial temperature smaller than T_* will not contain galaxies. When heating ceases at z_h, the gas cools adiabatically $T = T_*(1+z)^2(1+z_h)^{-2}$, allowing the "lighting" of haloes of smaller velocity. As the gas cools star formation takes place, and these haloes will then be identifiable as galaxies even if they will later be embedded into larger haloes. The mass function of "luminous" haloes can now be calculated in a straightforward way. The number of "lighted objects" of a given mass is given by the number of haloes of that mass at the redshift of first cooling for that mass. For a mass function

$$N(m)\, dm \;\propto\; (1+z)\, m^{-2}\, dm$$

with $L \propto \sigma^\beta$, and taking the relation $\sigma^2 \propto (1+z_f)\, m^{2/3}$, the luminosity function is calculated as

$$N(L)\, dL \;\propto\; L^{-\alpha}\, dL$$

with $\alpha = 1 + 0.5/\beta$. This slope agrees with observations. Note, however, that this argument holds only for a finite range that corresponds to $T_*/(1+z_h)^2 - T_*$. This implies a luminosity range of $(1 + z_h)^\beta$ and a turn-off is therefore expected in the luminosity function for luminosities fainter than M $\approx$ –12 to –15

To conclude, let us highlight some of the strengths and weaknesses of this model. We have been able to predict the right slope of the luminosity function, starting from the reliable hypothesis that the gas is hot. We have also predicted a cut-off or at least a turn-off in the faint-end of the luminosity function, as well as suggested that the IGM should be at a temperature of a few 10^4 K. The major problem is certainly the inclusion of heating and cooling details, in order to check whether the predicted evolution of

410

the LF is consistent with, for instance, the constraints given by deep counts of galaxies. Also, it remains to be seen whether the assumption of fragmentation –our main hypothesis– is sustained by numerical simulations.

References

Dekel, A., Silk, J., 1986, *Ap.J.* **303**, 39.
Newman, W.I., Wasserman, I., 1990, *Ap.J.* **354**, 411.
Peacock, J.A., Heavens,A.F., 1990, *M.N.R.A.S* **243**, 133.
Press, W.H., Schechter, P., 1974, *Ap.J.* **187**, 425.
Schaeffer, R., Silk, J., 1985, *Ap.J.* **292**, 319.

LARGE-SCALE MOTIONS IN THE UNIVERSE: A REVIEW

EDMUND BERTSCHINGER

Department of Physics, MIT, Cambridge, MA 02139 USA

ABSTRACT

Galaxy redshift and distance measurements may be combined to give deviations from uniform Hubble expansion called peculiar velocities. During the last several years, large numbers of galaxy distance measurements have been made for spiral and elliptical galaxies. These data have now been combined and synthesized to produce maps of the large-scale velocity, density and potential distributions. Because all matter, dark as well as luminous, generates a gravitational field, peculiar velocity surveys provide a measure of the matter distribution that cannot be obtained from galaxy redshift surveys alone. In combination with the IRAS galaxy redshift survey, the peculiar velocity field maps have been used to test the hypothesis that dark matter and galaxies are distributed similarly on large scales. This article reviews recent observational and theoretical advances in the study of large-scale motions.

1. INTRODUCTION

Measurements and analysis of large-scale motions have recently developed into a powerful new method for studying large-scale structure. Unlike galaxy redshift surveys, which map the large-scale distribution of luminous galaxies, surveys of galaxy peculiar motions (deviations from uniform Hubble expansion) allow the mapping of the large-scale distribution of all matter, dark and luminous. Although the number of galaxies with relatively accurate peculiar velocity measurements (with standard errors smaller than 25% of the distance) is still small (about 1000) compared with the number with measured redshifts, these data already allow us to draw significant conclusions about the large-scale distribution of matter in the universe.

There are at least three major theoretical motivations for undertaking detailed mapping of peculiar velocities. The first is to make a dynamical estimate of the cosmological density parameter Ω. For a given amplitude of density fluctuations, peculiar velocities increase with Ω (roughly as $\Omega^{4/7}$ if the perturbations are not strongly nonlinear; cf. Lightman and Schechter 1990). For a long time such estimates have been based on a single peculiar velocity only, that of the Local Group relative to the cosmic microwave background radiation (CMB) or to the Virgo cluster of galaxies (see Davis and Peebles 1983 for a review). Extending our knowledge of peculiar velocities beyond the Local Group offers an important consistency check on dynamical estimates of Ω and, more generally, on gravitational instability theory.

The second theoretical motivation for peculiar velocity studies is to test whether galaxies and dark matter are distributed similarly on large scales. In fact, dynamical estimates of Ω require assumptions about how galaxies trace mass, so this question is related empirically to the one above. There is reason to believe that on small scales (up to $\sim 10~h^{-1}$ Mpc) mass is more uniformly distributed than light (i.e., galaxies). (See Dekel and Rees 1987 for a review of this idea of biased galaxy formation.) The amount of bias, if any, prevaling on large scales (i.e., after spatially smoothing the matter and light distributions over scales $\gtrsim 10~h^{-1}$ Mpc) is presently unknown. Measuring the large-scale mass-to-light ratio is best done by mapping the large-scale velocity field.

Both of these first two goals require combining peculiar velocity data with complete galaxy redshift surveys. The latter are necessary for determining the galaxy density field. Peculiar velocity surveys, based on galaxy redshifts and distances, need not be complete because the galaxies are used only as tracers of the velocity field, much as stars or neutral hydrogen clouds are used as tracers of velocity fields within galaxies. Some coordination of redshift-distance surveys and complete redshift surveys is required to estimate Ω and to measure the large-scale mass-to-light ratio.

The third motivation for studying large-scale motions does not require a commensurate redshift survey: probing the initial conditions for the formation of structure. Up to now, most investigations of structure formation have relied on testing specific

theoretical models, e.g., the biased cold dark matter model (cf. Frenk *et al.* 1989 and references therein). However, the peculiar velocity field on large scales is expected to reflect directly the primordial fluctuations in the linear regime, with little or no correction needed for nonlinear evolution or biased galaxy formation. Maps of the initial peculiar velocity field, or of the derived density and potential fields, would provide an important guide for theories of dark matter and the origin of the initial fluctuations (e.g., the inflation paradigm). From these maps, or more directly from the data, it should be possible to measure the primordial power spectrum and to test whether the initial fluctuation field is a Gaussian random process.

These are bold goals. However, progress has already been made in each of these areas, as will be discussed in this article. Continued progress rests largely on the efforts of many observers.

The purpose of this article is to review recent observational and theoretical advances in the study of large-scale motions. This is not meant to be a comprehensive review; rather, it is a progress report with the aim of encouraging further work in this productive and important area. We begin with a review of techniques for measuring galaxy distances and a critical discussion of their errors. Then follows a discussion of recent observations and a comparison of velocity field measurements made by different techniques. These results and others are used to test our fundamental assumption that galaxy distance estimates measure peculiar velocities with reliable error estimates as opposed to measuring some peculiar properties of galaxies unrelated to their motions. We then discuss the potential flow reconstruction technique of Bertschinger and Dekel (1989) and review their recent results. A brief comparison is made with the IRAS redshift survey of Strauss and Davis (1988) and Yahil (1988), focusing on the questions of whether mass traces light on large scales and the value of Ω. After that we review several other analysis techniques and velocity-field statistics that have been applied to redshift-distance surveys.

For a comprehensive review of observations of large-scale motions up to mid-1989, see Burstein (1990). A thorough discussion of the Great Attractor (GA) models was given by Faber and Burstein (1988). Earlier reviews have been given by Dressler (1987*b*), Rubin (1989), and Gunn (1989).

2. GALAXY DISTANCE MEASUREMENT TECHNIQUES AND THEIR ERRORS

Two main techniques have been brought to bear for measuring distances of large numbers of galaxies: the Tully-Fisher (luminosity-line width) relation (Tully and Fisher 1977; Aaronson, Huchra, and Mould 1979; Pierce and Tully 1988) for spirals and the D_n-σ (diameter-velocity dispersion) relation (Dressler *et al.* 1987*a*; Djorgovski and Davis 1987; Dressler 1987) for ellipticals and S0s. Both of these techniques are based on empirical correlations between a distance-dependent quantity (flux or angular diameter) and a distance-independent quantity (circulation rotation speed or central velocity dispersion).

The theoretical basis for these correlations is not understood.

The measured scatter about the mean relations translates into rms errors in estimated distance of about 21% (0.41 magnitude) for the D_n-σ relation and 16% (0.32 magnitude) for the infrared or red Tully-Fisher relation (Burstein 1990). Some of this scatter is observational, arising primarily from photometric errors, implying a surprisingly tight correlation between the kinematic and photometric properties of galaxies. These correlations are an unexpected and unexplained gift, undoubtedly containing important clues about the process of galaxy formation. For our present purposes, we simply borrow them as tools for estimating distances, blissfully ignorant of their *raisons d'etre*.
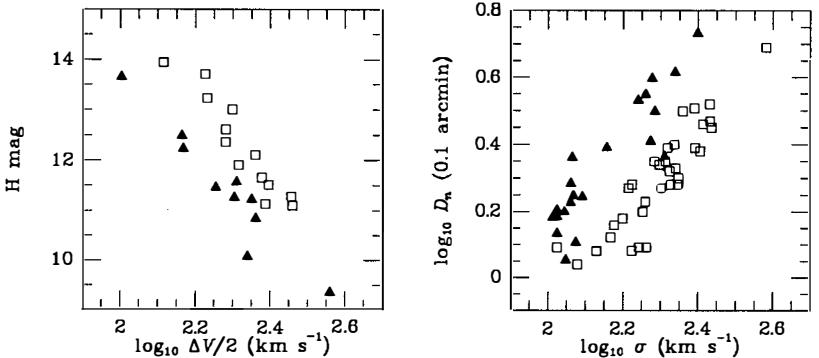
Fig. 1: Infrared Tully-Fisher (*left*) and D_n-σ (*right*) relations for the Coma (*squares*) and Hydra (*triangles*) clusters.

Figure 1 shows the empirical infrared Tully-Fisher (IRTF) and D_n-σ relations for galaxies in the Coma and Hydra clusters. Coma galaxies are smaller and fainter than Hydra galaxies because they are more distant. The data are taken from Aaronson *et al.* (1986) and Faber *et al.* (1989), as compiled by Burstein (1989). The relative Coma-Hydra distance modulus derived from the offset between the distance-indicator relations for the two clusters is 1.17 magnitudes from the D_n-σ relation and 1.31 magnitudes from the IRTF relation. The ratio of redshifts in the CMB frame gives a relative distance modulus of 1.25 magnitudes. Thus, from the D_n-σ distances we would infer that Hydra has an inward peculiar velocity relative to Coma (the ratio of redshifts is greater than the ratio of distances), while the IRTF distances lead to the opposite conclusion.

Note that peculiar velocities can be inferred without knowledge of the absolute distance scale — the value of the Hubble constant need not be known. To see this more clearly, note that the radial peculiar velocity is

$$v_r = cz - H_0 r \ . \tag{1}$$

For a galaxy with large distance r_1, whose peculiar velocity is much less than its redshift

z_1, we have $H_0 r_1 \approx c z_1$. If other galaxy distances are measured relative to this one, we obtain the distance in units of $\mathrm{km\,s^{-1}}$: $H_0 r = (r/r_1) c z_1$. Accordingly, we hereafter measure all distances and velocities in units of $\mathrm{km\,s^{-1}}$. In practice it is better to define the relative distance scale using a large number of galaxies at large distances, in order to reduce the effects of both residual peculiar velocities and distance errors. In this way, Burstein (1989) has determined the radial peculiar velocities of the Coma and Hydra clusters to be $-179 \pm 270 \ \mathrm{km\,s^{-1}}$ and $-268 \pm 197 \ \mathrm{km\,s^{-1}}$, respectively, from the D_n-σ distances, and $-179 \pm 328 \ \mathrm{km\,s^{-1}}$ and $30 \pm 204 \ \mathrm{km\,s^{-1}}$ from the IRTF distances. These numbers are consistent with our conclusions about the relative peculiar velocities reached above, but we note that the uncertainties, even for these well-sampled clusters, are large.

Figure 1 and the arguments presented above make it seem straightforward to measure peculiar velocities. However, they also suggest two important reasons for caution: random errors and systematic errors. Prudence is mandated by our very lack of understanding of the Tully-Fisher and D_n-σ relations, so understanding the effects of errors requires careful analysis of the empirical correlations. One should not take the accuracy of the derived peculiar velocities for granted.

First, let us consider the possible systematic errors: Granted, there appears to be a good correlation of internal velocity with diameter and/or luminosity in each cluster. But, how do we know that galaxies of the same circular velocity have the same absolute magnitude in Virgo and Coma? Perhaps, as suggested by Djorgovski *et al.* (1989) and Silk (1989), the zero-point of the D_n-σ or Tully-Fisher relation varies with environment. It is plausible, for example, that galaxies of the same velocity dispersion in a high-density region might be more luminous than those in the field. This would translate into artificially inferred outward motions for galaxies in dense regions, such as the Hydra and Centaurus clusters, which are, in fact, inferred to be falling toward the Great Attractor. Spurious infall is a worrisome possibility. However, careful searches for environmental correlations in the D_n-σ relation (Giuricin *et al.* 1989; Burstein *et al.* 1990) have found no such effects at the level of a few percent, low enough that residual systematic errors are presently of little concern. Similar conclusions hold for the Tully-Fisher relation (Mould *et al.* 1989; Burstein 1990). Also, Donnelly *et al.* (1990) have shown that the L_X-L_{opt} correlation of ellipticals is improved when D_n-σ distances are used. Moreover, the peculiar velocity fields inferred separately from the D_n-σ and Tully-Fisher relations are in fairly good agreement (Faber and Burstein 1988; Burstein *et al.* 1990; Fig. 2 below), an unlikely outcome if both are affected by systematic errors (Gunn 1989; but see Silk 1989 for a different view). Finally, the consistency between the mass density field obtained below from peculiar velocities with the galaxy density field of the IRAS redshift survey would require systematic errors to accurately mimic the expected physical peculiar velocities, an implausible result. Although systematic zero-point errors are undoubtedly present at some level, it appears that they may be too small to perturb noticeably the inferred motions.

In contrast, we know that nonnegligible random errors are present in every galaxy distance measurement. They arise simply from the imperfect correlation of internal velocity with luminosity or size. Given the template D_n-σ or luminosity-line width relation obtained from calibrating cluster galaxies at a common distance, the best one can do for a single measurement is to place the galaxy on the mean regression line of the relation, introducing a random error drawn from the distribution of residuals from the mean relation. The standard error per galaxy, about 20% of the distance, exceeds typical peculiar velocities at distances beyond 3000 km s^{-1}. Without a distance indicator of smaller scatter, one can improve the signal-to-noise ratio only by measuring more galaxy distances. See Kraan-Korteweg *et al.* (1988), Pierce and Tully (1988), Faber and Burstein (1988), and Burstein and Raychaudhury (1989) for varying views on the magnitude of the scatter in the Tully-Fisher relation.

Potentially more troublesome is a systematic error traceable to the finite scatter of the distance-indicator relations: so-called Malmquist bias. This bias occurs only when peculiar velocities are plotted versus estimated distance (Lynden-Bell *et al.* 1988). It arises because there are more galaxies at large distance and in denser regions, giving a factor $r^2 n(\vec{r})$ that should multiply the probability distribution of true distance for given measurements, where n is the galaxy density. Neglecting this factor causes a systematic underestimate of galaxy distances by an amount proportional to $r\Delta^2$, where Δ is the standard error of $\ln(r)$. Lynden-Bell *et al.* (1988) show how to remove this bias for a uniform density of galaxies; this correction is applied to most distance samples in widespread use. Because the correction is quadratic in the random error, and is significant at large distances, it is important to know the random errors well and to try to minimize them.

Even with the correct homogeneous Malmquist correction, however, one is still subject to Malmquist bias arising from density gradients. For clusters of galaxies this bias may be reduced by averaging the distance measurements and applying a reduced Malmquist correction (Lynden-Bell *et al.* 1988), but it is more difficult to deal with large-scale density gradients. Removing the residual bias requires knowing the galaxy density field, which requires a complete sample obtained with a well-defined selection function. In practice, this bias may not be as worrisome as feared. Using Monte Carlo simulations, Dekel *et al.* (1990) estimated the systematic error in the large-scale velocity field (spatially smoothed with a Gaussian filter of radius 1200 km s^{-1}) due to inhomogeneous Malmquist bias to be less than 200 km s^{-1} out to a distance of 6000 km s^{-1}. This is smaller than the random errors in the smoothed velocity field and it has very little effect on any conclusions drawn from the data aside from the exact location of the GA infall center (cf. Burstein *et al.* 1990).

There are alternative methods for estimating distances from the same data that do not suffer from Malmquist bias. However, these methods generally require knowing the sample selection function or having a good model for the peculiar velocity field, without which they, too, suffer from biases. For detailed discussions of errors see Fall and Jones

(1976), Schechter (1977, 1980), Aaronson *et al.* (1982*b*, 1986), Teerikorpi (1987), Bottinelli *et al.* (1988), Lynden-Bell *et al.* (1988), Faber and Burstein (1988), Bothun *et al.* (1990), Burstein *et al.* (1990), and Dekel *et al.* (1990).

Distance errors are perhaps the most obvious and easiest to quantify of the several sources of noise in estimates of the large-scale velocity field. However, as Dekel *et al.* (1990) show, sparse and nonuniform sampling cause problems even in the absence of distance errors. For example, if there is a steep decline in the number density of galaxies, owing to a steep gradient in the selection function (e.g., across the Galactic plane), and if the galaxies are given equal weight, the smoothed velocity estimated in the low-density region will be contaminated by the velocity field from the densely sampled region, an effect Dekel *et al.* call sampling gradient bias. This bias, like Malmquist bias, can be corrected for if the galaxy sample is complete with a well-defined selection function; in practice the samples in use are too heterogeneous to allow a simple correction. Sampling gradient bias can also be decreased by decreasing the amount of smoothing applied to the data. Unfortunately, a large amount of smoothing is required with present samples in order to reduce the variance caused by random distance errors. The sampling gradient bias and the related sampling variance (caused by Poisson sampling of a velocity field with gradients inside the smoothing window) are particularly worrisome because they depend on the underlying velocity field that one is trying to determine. In practice, the best compromise seems to be to weigh each peculiar velocity measurement inversely by the sampling density, to use a moderately large amount of smoothing to reduce the effects of measurement errors, and then to exclude from further analysis poorly sampled regions. It should be noted that these and other biases are not unique to the particular analysis method used by Dekel *et al.* , but are generic to *all* analyses of large-scale peculiar velocities, including those without explicit smoothing.

In short, workers desiring to draw cosmological conclusions from galaxy redshift and distance measurements need to analyze carefully the effects of systematic and random errors. These errors become less troubling as the number of distance measurements increases. More accurate distance indicators would also help. In this regard, we look forward to the development of the luminosity fluctuation method (Tonry *et al.* 1989), which may yield substantially more accurate distances for elliptical galaxies within ~ 5000 km s^{-1} distance from the Local Group.

To conclude this section, we present a comparison of the radial velocity field implied by ellipticals and S0s with that from D_n-σ distances (from Faber *et al.* 1989; Lucey and Carter 1988; and Dressler and Faber 1990*a*) with spirals with IRTF distances (from Aaronson *et al.* 1982*a*, 1986, 1989; and Bothun *et al.* 1984). The treatment of the data is as was described by Bertschinger *et al.* (1990). We combine the individual galaxy and group peculiar velocities, using a tensor window function, into a smooth radial velocity field $v_r(\vec{r})$. The distance r is, of course, the estimated distance (with homogeneous Malmquist correction) and not the true distance, which is unknown. The Gaussian smoothing radius is

$1200 \, \mathrm{km \, s^{-1}}$. The radial velocity field is sampled on a Cartesian grid of spacing $1000 \, \mathrm{km \, s^{-1}}$, but only those grid points within $4000 \, \mathrm{km \, s^{-1}}$ distance from the center and with at least 4 distance measurements from both samples within a distance of $1500 \, \mathrm{km \, s^{-1}}$ from the given grid point, are retained. In other words, the comparison is made over a relatively small volume and only at points that are well sampled by both types of galaxies. This procedure is carried out separately for both the E/S0 and the S samples.
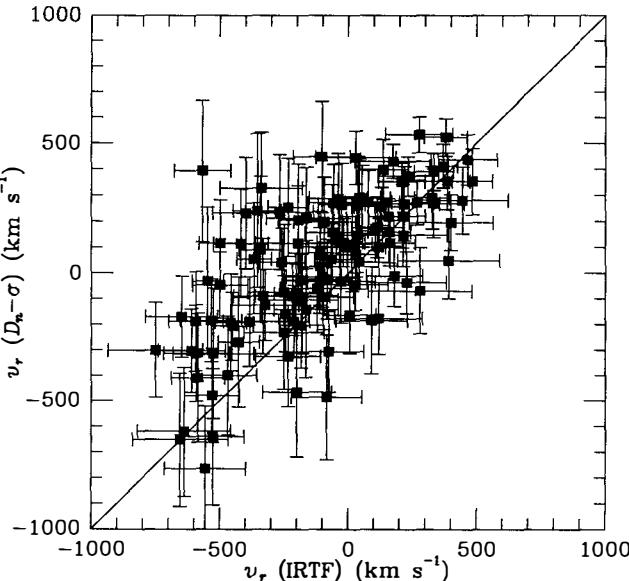
Fig. 2: Smoothed peculiar radial velocities from D_n-σ distance measurements versus those from IRTF distances.

The D_n-σ and IRTF radial velocities are plotted against each other in Figure 2. Error bars have been computed by making 100 noisy realizations of each data set as described in Bertschinger et al. We see that there is general agreement, with scatter, between the Tully-Fisher and D_n-σ peculiar velocity fields. There does appear to be a slight systematic offset, with the peculiar velocities of early-type galaxies being more positive than those of spirals. This also shows up as systematically larger redshifts for the E/S0s, at the same estimated distance, as for the spirals. Goodness of fit was tested by performing a linear regression of the data and computing the regression coefficients, the linear correlation coefficient and chi squared. Monte Carlo simulations were then performed to test the hypothesis that the measured D_n-σ and IRTF velocity fields are noisy versions of the same underlying velocity field. These simulations began by replacing the measured peculiar velocities with a model velocity field (the Faber and Burstein 1988 GA+Virgo model), so that the underlying D_n-σ

and IRTF peculiar velocity fields were identical. Noise was then added to each, and 100 realizations were made. The same statistics computed from the data were computed from each realization, yielding the probability distribution of each statistic. The results indicate that the measured value of chi squared shows consistency between the two velocity fields, but that the slope and offset are significantly different (at about 95% confidence) from what one would expect if the underlying velocity fields really are identical. The implication is that the estimated random errors of the D_n-σ and IRTF relations are about right, but that the velocity field inferred with ellipticals is noticeably different from that of the spirals. The simplest explanation is that there is an error in the relative zero-point calibration of the two relations. If the distances of nearby spirals are decreased by 4% (0.08 magnitudes in distance modulus), the offset is removed. Because the distances of nearby spirals have been calculated by Burstein (1989) using a different Tully-Fisher relation than that used for the cluster spirals, an error of this magnitude is not unreasonable. With this correction, the peculiar velocity fields implied by nearby spirals and ellipticals are consistent within the measurement errors.

3. RECENT OBSERVATIONAL RESULTS

Before summarizing new results published during the last year, it is useful to summarize the observational status of the subject of large-scale motions as it was at the beginning of 1989. At that time, the main question being asked was, "Is the Great Attractor real?" Lynden-Bell *et al.* (1988) had published a detailed analysis of their D_n-σ measurements of distances to about 400 elliptical galaxies, showing that the peculiar velocity field within about 6000 km s^{-1} distance from the Local Group was fit well by infall toward a point beyond the Hydra-Centaurus supercluster, at a distance of 4350 ± 350 km s^{-1}. The magnitude and coherence of this flow is larger than was expected by practically everyone (although it is comparable with the original Rubin *et al.* 1976 result) and it poses a strong challenge to the popular models of galaxy and structure formation (Bertschinger and Juszkiewicz 1988). Given the thoroughness of the "Seven Samurai" (7S) in their series of papers (see Faber *et al.* 1989 and references therein), it was hard to find a flaw in their result, but most astronomers were skeptical.

If the results of the 7S were correct, then it seemed difficult to reconcile the large-scale peculiar velocity field with the large-scale gravitational field inferred by Yahil (1988) from the IRAS redshift survey (Strauss and Davis 1988). This redshift survey is complete over about 90% of the sky to a given 60μ flux limit and is based on galaxy identifications from the IRAS point source catalog. It is superior to optically-selected samples in its being little affected by Galactic dust extinction. The gravitational field predicted from the observed distribution of IRAS galaxies contains two equally prominent peaks on the scale of the peculiar velocity surveys, the GA peak in the direction of Hydra-Centaurus, and the Perseus-Pisces supercluster at about the same distance on the opposite side of the sky. On

large scales, the peculiar velocity field is expected to be proportional to the gravitational field. As Yahil (1988) and Strauss (1989) showed by comparing the gravitational field with the radial velocity field of 7S, there was no evidence from the galaxy distance measurements existing then for any significant inflow into the Perseus-Pisces supercluster. This result was confirmed by Dekel and Bertschinger (1990) by comparing the mass density field with the galaxy density field. The second main question being asked was therefore, "What about the Perseus-Pisces supercluster?"

Newer results have clarified these issues somewhat. First, the reality of the GA was bolstered by the demonstration by Faber and Burstein (1988) of the agreement of the nearby velocity field implied by D_n-σ distances of early-type galaxies with that obtained from infrared Tully-Fisher distances of spirals (Aaronson et al. 1982a; Bothun et al. 1984; Aaronson et al. 1986). This agreement was strengthened by Aaronson et al. (1989), who measured spiral distances to six clusters in the Hydra-Centaurus supercluster and found large outward peculiar velocities for three of them. Staveley-Smith and Davies (1989) also found evidence for a dipolar and quadrupolar flow directed toward the GA. This latter study disagreed with Lynden-Bell et al. about the distance to the GA infall center, preferring a more nearby center coincident, perhaps, with the Centaurus clusters themselves. However, it was widely recognized that the flow in the Hydra-Centaurus supercluster itself is complex and nonlinear (e.g., Lucey et al. 1986), making it seem plausible that the spiral and elliptical results were not inconsistent; perhaps the spiral-rich groups do not coincide with the elliptical-rich cluster cores. Lucey and Carter (1988) also found large outward peculiar velocities in the Hydra-Centaurus supercluster using D_n-σ distances.

Another major step forward came with the completion of a new redshift survey covering one steradian around the direction of the GA by Dressler (1988) and the subsequent measurement of D_n-σ and Tully-Fisher distances to more than 200 of these galaxies by Dressler and Faber (1990a,b). These data confirmed the original 7S determination that the foreground Hydra-Centaurus groups have larger outward peculiar velocities. They also showed evidence that galaxies beyond the infall center, with peculiar velocities directed toward us, have finally been discovered. Although the exact location of this infall center is dependent on somewhat uncertain Malmquist corrections, its existence now seems secure (Burstein et al. 1990; Dressler and Faber 1990b). Dekel et al. (1990) and Bertschinger et al. (1990) concluded from their analysis of the effects of distance errors that residual Malmquist bias produces at most a 200 km s^{-1} error in the velocity field near the GA smoothed with a Gaussian of radius 1200 km s^{-1}, compared with maximum smoothed infall velocities exceeding 600 km s^{-1}. Although Malmquist effects are likely to be larger in the unsmoothed redshift-distance maps of Dressler and Faber, the relatively small scatter in these maps, especially those with Tully-Fisher distances, argues for the reality of the "S-wave" perturbation of the Hubble flow.

A spanner was almost thrown into the works with the rediscovery, by Scaramella *et al.* (1989), of an enormous supercluster beyond the GA at a redshift of about 14000 km s^{-1}. This concentration of rich clusters had been noticed earlier by Shapley (1930). If these clusters are as prominent in the dark matter distribution as they are in the galaxy distribution, they may perturb significantly the flow in the Hydra-Centaurus supercluster (Raychaudhury 1989). Perhaps they do; that would not obviously contradict any observations at present. However, one should be cautious in estimating the gravitational effect of distant clusters given the lack of a complete survey. The ROSAT X-ray satellite will help in this regard, by providing a uniform sample of X-ray emitting clusters. It should also be noted that a void between the foreground Hydra-Centaurus supercluster and the background "Attratore di Tutti Attratore" (Dekel 1988) would diminish the predicted perturbations. The Shapley concentration very likely has little effect on the motion of the Local Group (Raychaudhury 1989).

And what about Perseus-Pisces? This question is still not answered satisfactorily, although some progress has been made. Dekel *et al.* (1990) and Bertschinger *et al.* (1990) pointed out that the peculiar velocity field is very poorly sampled in the Perseus-Pisces supercluster. The velocity and inferred mass density fields in this region are strongly affected by the peculiar velocity of a single object, the Perseus cluster. Bertschinger *et al.* noted that when one velocity measurement has the dominant weight in a region, the fitted velocity field is necessarily smoother than the true one (since one velocity determines only a bulk flow), so that the inferred mass density perturbation (proportional to $-\vec{\nabla} \cdot \vec{v}$ in the linear regime) is biased to lower amplitude, an unfortunate but unavoidable consequence of the sampling gradient bias discussed above. They concluded that more distance measurements in the Perseus-Pisces region are needed before it can be determined that there is a convincing discrepancy between the large-scale peculiar velocity field and the galaxy distribution.

Fortunately, Haynes and Giovanelli (1988) have been undertaking a large survey of spiral galaxies in the Perseus-Pisces supercluster using the Arecibo radiotelescope to measure 21 cm redshifts and linewidths. The completion of this redshift survey, together with the photometry necessary for Tully-Fisher distances, will lead to a dramatic increase in the sampling of the velocity field in the hemisphere opposite to the GA. Many Tully-Fisher distances have already been obtained by Willick (1990), who concludes that the groups in the direction of Pisces are falling in toward the Local Group and hence toward the GA with velocities comparable to those measured by Lynden-Bell *et al.* (1988) (several hundred km s^{-1}). In addition, the data show some evidence for small-scale compressional motions toward the Pisces cluster. Although the agreement with 7S is encouraging, these results still do not reconcile the apparent disagreement with the IRAS redshift survey. Understanding the dynamics of the Perseus-Pisces supercluster is one of the major challenges presently facing workers in this field.

4. POTENTIAL FLOW RECONSTRUCTION

With the accumulation of large numbers of galaxy distance measurements during the last few years, the need has arisen for a general, model-independent method for analyzing these data to extract the large-scale peculiar velocity field. Such a method, based on the assumption of potential flow, was developed by Bertschinger and Dekel (1989). This method, nicknamed POTENT, extracts the smoothed three-dimensional velocity field from samples of galaxy distances and redshifts. In addition, assuming that gravity is responsible for generating the peculiar motions, the mass density and gravitational potential fields are extracted. All of this is possible from samples of galaxy distances and redshifts. How?

A simple counting argument suggests that it should be possible to reconstruct the smoothed three-dimensional velocity and density fields from measurements of the radial peculiar velocity only. According to gravitational instability theory, the initial conditions for the formation of large-scale structure were determined by one scalar field, the density fluctuation field (or, equivalently, the gravitational potential). Radial peculiar velocity measurements provide one constraint at each sampled point, implying that it is possible, in principle, to recover the initial conditions and the present three-dimensional fields.

The practical implementation of this idea is based on the fact that the peculiar velocity field is predicted by gravitational instability theory to be irrotational (curl-free) outside of regions of orbit-crossing and dissipation, with $\vec{\nabla} \times \vec{v} = 0$. (This is not in conflict with the spinning up of galaxies by tidal torques, a process that is dependent upon orbit-crossing on galactic scales; cf. Peebles 1969.) For an irrotational (i.e., potential) flow, the peculiar velocity field is derived from a single scalar velocity potential Φ:

$$\vec{v} = -\vec{\nabla}\Phi \ . \tag{2}$$

If potential flow prevails on the large scales of interest, then the tangential components of the peculiar velocity field can be obtained using equation (2), with the potential determined by integrating the radial component along radial rays:

$$\Phi(r, \theta, \phi) = -\int_0^r v_r(r', \theta, \phi)\, dr' \ . \tag{3}$$

Before equations (2) and (3) can be applied, the peculiar velocities of individual galaxies must first be interpolated and smoothed to provide an estimate of $v_r(\vec{r})$ throughout the volume of interest. It is desirable that the smoothing procedure approximate as closely as possible a convolution of the underlying velocity field by a uniform "window function", e.g., a spherical Gaussian filter. A tensor window function is necessary because the smoothing averages measurements made in different directions. In theory, enough smoothing is required to filter out any vorticity created by nonlinear processes on small scales. In practice, the peculiar velocity field is traced so sparsely by existing samples that a large degree of smoothing is required simply to define a radial velocity field with a useful signal-to-noise

ratio. Because of the problems introduced by random distance errors, Malmquist bias, sampling gradient bias and sampling variance discussed above, the practical implementation of the tensor window function smoothing requires a complicated set of compromises. A thorough discussion of these issues has been presented by Dekel *et al.* (1990).

Given the smoothed radial velocity field, it is straightforward to integrate equation (2) numerically on a grid in spherical coordinates. One recent change made in the POTENT algorithm is that we now interpolate the velocity potential to a cubic grid before differentiation rather than interpolating later. With sufficiently fine spherical and cubic grids, this new procedure yields a velocity field virtually identical to the one obtained using the method described by Dekel *et al.* (1990), but it makes the density reconstruction simpler. Dekel *et al.* tested the potential, velocity and density reconstructions extensively in the presence of distance errors and sparse sampling using Monte Carlo simulations based on an N-body simulation. The outcome of their tests is that they have developed a reliable way of estimating and/or controlling random errors and known systematic errors (Malmquist bias and sampling gradient bias) to the extect that they affect the reconstructions.

Bertschinger *et al.* (1990) applied POTENT to a set of nearly 1000 galaxy distance measurements, 56% of them being D_n-σ measurements by 7S, Lucey and Carter (1988) and Dressler and Faber (1990a), and the rest IRTF distances from Aaronson *et al.* (1982a, 1986, 1989) and Bothun *et al.* (1984). These are the same samples compared in Figure 2. They provide maps of the reconstructed velocity potential, the velocity field, and the mass density field, along with a few global statistics of the velocity field. Their results largely confirm the Great Attractor picture of Lynden-Bell *et al.* (1988). The velocity potential within a distance of 6000 km s^{-1} from the Local Group is found to be dominated by the GA, with the difference in potential between the GA and the Local Group exceeding 2×10^6 (km s^{-1})^2, corresponding to a mean flow amplitude of about 500 km s^{-1} along the line of sight toward the GA. The peak density contrast in the GA, with an effective Gaussian smoothing radius of 1400 km s^{-1}, is given by $\Omega^{4/7}\delta\rho/\rho = 1.2 \pm 0.4$. As noted above, it is necessary to assume a value of Ω in order to relate peculiar velocity to density (and vice versa).
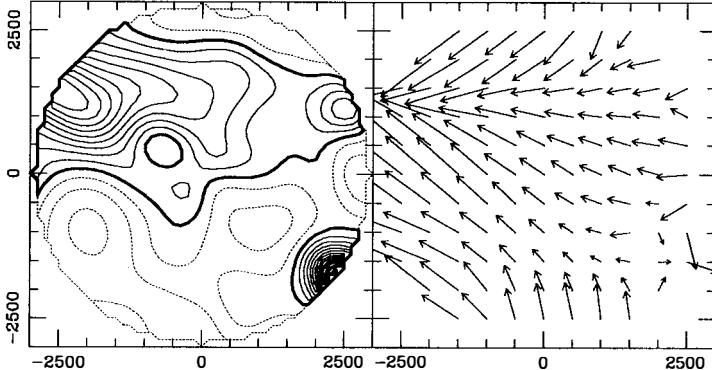
Fig. 3: POTENT reconstruction of the density and velocity fields
of the Local Supercluster.

Figure 3 shows a high-resolution view of the density and velocity fields in the
Supergalactic plane within 3000 $km\,s^{-1}$ (from Bertschinger *et al.* 1990). The contour
spacing is 0.2 in $\delta\rho/\rho$, with negative contours dotted and the zero contour heavy. The
effective Gaussian smoothing radius is 500 $km\,s^{-1}$ for the velocity and 700 $km\,s^{-1}$ for
the density. The ridge of the Local Supercluster is evident as the flattened structure at
$Y \approx 1500$ $km\,s^{-1}$, extending from Centaurus on the left, through the Virgo Southern
Extension, and to the Ursa Major cluster. The two density peaks at $X \approx 2500$ $km\,s^{-1}$
are probably noise artifacts, arising from too few data. Comparison with the maps of the
galaxy density in the *Nearby Galaxies Atlas* (Tully and Fisher 1987) reveals an encouraging
correlation between the distributions of light and mass.

5. COMPARISON WITH THE IRAS REDSHIFT SURVEY

The peculiar velocity field may be compared with the galaxy density field in order
to test whether mass is distributed like galaxies and, if this appears to be true, to make a
dynamical estimate of Ω. As we discussed in Section 1, this requires both a peculiar velocity
survey and a complete galaxy redshift survey. At present, the best all-sky redshift survey
for this type of comparison is the IRAS redshift survey of Strauss, Davis and Yahil (Strauss
and Davis 1988, 1990; Yahil 1988; Strauss 1989). The survey covers 88% of the sky, with
more than 2500 redshifts of galaxies with 60 micron flux greater than 1.9 Jy. The galaxies
were selected from the IRAS point source catalog using color criteria; they are almost
all late-type spirals, since the IRAS satellite was sensitive to the infrared re-radiation of
starlight absorbed by interstellar dust grains. The primary advantage of an infrared-selected
redshift survey over optical surveys is the much better sky coverage, owing to the negligible

Galactic extinction at a wavelength of 60 microns. Optical redshift surveys become seriously incomplete at Galactic latitudes $|b| \lesssim 20°$, while the IRAS survey is essentially complete for $|b| > 5°$.

Given measurements of the peculiar velocity field and of the galaxy density field, there are two ways to make the comparison of mass and light. The first way is to estimate the gravitational field from the galaxy density field assuming that galaxies trace the mass and then to predict the peculiar velocity field on the basis of linear or quasi-linear theory. The second method, requiring POTENT, is to predict the mass density field from measured peculiar velocities and then to compare with the galaxy density field. Each method has advantages and disadvantages.

First, let us review the velocity field comparison. In practice, the predicted peculiar velocities are used to correct the galaxy redshifts to predicted distances, so that the velocity field comparison can be made in position space instead of redshift space. The radial components of the predicted peculiar velocity field are then compared directly with the measured radial peculiar velocity field. This technique has been pursued extensively by Yahil (1988, 1990a) and Strauss (1989). The comparisons show some qualitative agreement between mass and light but with lots of scatter and some systematic differences. The most severe discrepancy, as noted in Section 3, appears in the relative importance of the GA and the Perseus-Pisces supercluster. The gravity field obtained from the IRAS redshift survey predicts that the Perseus-Pisces supercluster should have an effect on the local flows comparable to that of the GA, while the measured peculiar velocity field shows very little evidence for any effect of Perseus-Pisces.

The velocity field comparison has the advantage that it works directly with the measured radial peculiar velocities, with minimal massaging of these data. Three-dimensional peculiar velocities reconstructed from POTENT could be used equally well, but the only additional advantage would be the possibility of applying a relatively uniform spatial smoothing. Comparison of the mass density field reconstructed from POTENT with the IRAS galaxy distribution suffers from the fact that the mass density field is reconstructed relatively less accurately than the velocity field. This is because the mass density field depends, in linear theory, on the gradient of the peculiar velocity field:

$$f(\Omega)\delta\rho/\rho \approx -\vec{\nabla} \cdot \vec{v} , \qquad (4)$$

where $f(\Omega) \approx \Omega^{4/7}$ (Lightman and Schechter 1990). Differentiating a noisy field makes the noise worse. As Dekel et al. (1990) show, the mass density field reconstructed from POTENT has a smaller signal-to-noise ratio than the velocity field, which is still noisier than the potential (since the latter integrates the noisy velocity field).

While POTENT is stronger with the peculiar velocity field, the peculiar velocity field predicted from the IRAS redshift survey is less well determined than the galaxy density itself. The reason is two-fold. First, in order to determine the gravity field one needs, in principle,

to sum over all the mass (or galaxies) over a very large volume. Second, it is easier to make a quasi-linear estimate of the density from the velocity than vice versa. The first problem arises in practice in two ways: first, the IRAS sample is cut at a distance of 8000 km s^{-1} from the Local Group, so that the gravitational effect of mass (and voids) beyond this distance is totally absent; and, second, even within this volume the sampling of IRAS galaxies is relatively sparse, so that Poisson sampling fluctuations (shot noise) are nonnegligible. Because gravity is long-ranged, errors in the mass distribution at large distance affect the predicted velocity field throughout the sample. These are difficult problems, whose assessment requires detailed Monte Carlo simulations applying the IRAS selection effects and estimation methods to known models. The realization of the seriousness of these problems has led Yahil (1990a) to avoid making the velocity field comparison in the CMB frame, since it now appears that the IRAS survey may not trace the galaxy distribution over a large enough volume to allow the dipole motion of the Local Group to be predicted to better than 200 km s^{-1} or so. Once the dipole is fixed, higher-order components of the velocity field should be predicted more accurately, at least within a distance of 4000 km s^{-1} of the Local Group.

Although the POTENT density reconstruction has smaller signal-to-noise ratio than the measured velocity field, this disadvantage may be compensated for by the fact that the galaxy density field, as traced by the IRAS redshift survey, is known rather well. Moreover, the density field comparison is appealing because it is purely local, so that what you see is what you get. One can test whether mass traces light point-by-point throughout the volume surveyed. Moreover, although the errors of the POTENT density reconstruction are relatively large, they are reasonably well understood from Monte Carlo simulations (Dekel *et al.* 1990, Bertschinger *et al.* 1990).
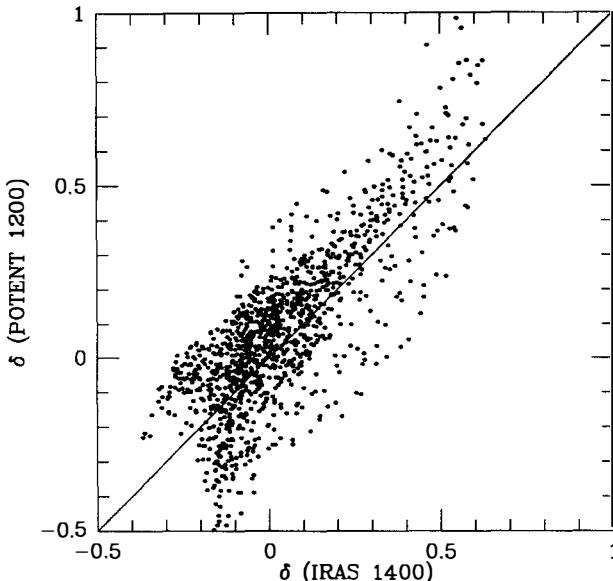
Fig. 4: Smoothed mass density contrast (POTENT) versus
the smoothed galaxy density contrast(IRAS).

A comparison of the IRAS and POTENT density fields within a distance of about 4000 km s^{-1} from the Local Group is presented in Figure 4. With a conservatively large amount of smoothing, defined by a Gaussian filter of radius 1200 km s^{-1}, the galaxy density field and the mass density field agree remarkably well under the assumptions that $\Omega = 1$ and that the IRAS galaxies are an unbiased tracer of mass. Quantitative tests and more details are provided by Yahil (1990b). This comparison specifically excludes the region around Perseus, since there are too few tracers of the peculiar velocity field there to allow a determination of its divergence. Preliminary Monte Carlo simulations indicate that the differences between the IRAS and POTENT density fields are consistent with the effects of distance errors in POTENT and shot noise in IRAS, and that the relative amplitudes of the two density fields agree to within ±25% (90% confidence). The results certainly do not prove the hypotheses that $\Omega = 1$ and that IRAS galaxies exactly trace mass on large scales, but the overall agreement encourages one to think that the goals outlined at the beginning of this article may be realistic. However, it is important to be cautious at this early stage. The disagreement between the predicted velocity field and the measured one, especially around the Perseus-Pisces supercluster, are evidence enough that we should not have confidence that our results are as significant as the formal errors suggest.

6. OTHER ANALYSIS TOOLS

As was emphasized in Section 1, the large-scale peculiar velocity field is ideal for computing statistics of the large-scale matter distribution that theorists can use for testing models. Not only are we free (one hopes!) from the usual assumptions about how mass traces light, but linear theory is generally adequate on large scales, implying that simple comparisons need not be complicated by nonlinear evolution. There are several statistics that have been applied directly to the radial peculiar velocity measurements, and others that may be applied to the three-dimensional velocity, potential and density fields reconstructed by POTENT. A brief sampling of recent results is presented here.

The first statistic to be computed from the 7S survey was the "bulk flow" or mean velocity averaged over a sphere of radius 6000 $km\,s^{-1}$ (Dressler *et al.* 1987*b*). The original value (599 ± 104 $km\,s^{-1}$ toward Supergalactic latitude and longitude $L = 173°$, $B = -5°$) has decreased with the addition of new data; also, Kaiser (1988) and Regős and Szalay (1989) showed that the effective window function volume was significantly smaller than a tophat of radius 6000 $km\,s^{-1}$. The most recent estimates of bulk flow, based on volume-weighted averages of the three-dimensional velocity field, were presented by Bertschinger *et al.* (1990). They measured the average velocities in spheres of radii 4000 and 6000 $km\,s^{-1}$ centered on the Local Group (with prior, additional smoothing by a Gaussian of radius 1200 $km\,s^{-1}$) to be 388 ± 67 $km\,s^{-1}$ toward $L = 177°$, $B = -15°$ and 327 ± 82 $km\,s^{-1}$ toward $L = 194°$, $B = 5°$, respectively, compared with predicted rms velocities of 287 and 224 $km\,s^{-1}$ for the unbiased cold dark matter model. Although the discrepancy with the cold dark matter model is less severe than was thought originally (e.g., Bertschinger and Juszkiewicz 1988), it is still serious for the highly biased $b = 2.5$ model.

Another statistic that measures the coherence and amplitude of the large-scale motions is the velocity correlation function, $\xi_v(r) = \langle \vec{v}(0) \cdot \vec{v}(\vec{r}) \rangle$, where the average is taken over pairs of points separated by a distance r. Estimating this from radial peculiar velocities is somewhat tricky. Górski *et al.* (1989) and Groth *et al.* (1989) used different estimators but reached similar conclusions: the peculiar velocity field traced by ellipticals with D_n-σ distances measurements has a large amplitude and is coherent over scales of at least 20 h^{-1} Mpc. They also conclude that the velocity field of spirals (from IRTF measurements) has smaller amplitude, but one should note that the sample volumes are not compatible, since the spiral sample densely samples the Local Supercluster within a distance of 2500 $km\,s^{-1}$ and is very sparse beyond, while the elliptical sample is more uniform out to a distance of 6000 $km\,s^{-1}$.

Ostriker and Suto (1989) introduced a new statistic called the cosmic Mach number $M = V/\sigma$, the ratio of bulk-flow velocity V in a volume to the velocity dispersion σ about that bulk flow within the same volume. This statistic has the very nice feature that it is independent, in the linear regime, of the amplitude of the primordial power spectrum, depending only on its shape. Thus, it should allow a nice test of the primordial Zel'dovich

spectrum. Ostriker and Suto estimate the Mach number from the data on large-scale motions and they conclude that it is so large as to rule out the Zel'dovich spectrum with $\Omega = 1$ and cold or hot dark matter, independently of the amplitude or the linear bias factor b. However, they caution that their observational estimates must be considered tentative. Since they did not estimate V and σ in the same volume, and they did not perform a detailed error analysis, their estimates should not be used to reject theories at this stage.

Juszkiewicz *et al.* (1990) have introduced another statistic that is sensitive to power on large scales, the potential structure function $g(r) = \langle (\Phi(\vec{r}) - \Phi(0))^2 \rangle$. This statistic has the advantage that it is fairly insensitive to random peculiar velocity errors. Juszkiewicz *et al.* test various scenarios against the measurements derived from POTENT. They conclude, based on Monte Carlo simulations of both the observations and the theory, that the $b = 2.5$ cold dark matter model is rejected at the 95% confidence level but that the unbiased amplitude cannot be excluded, based on this test.

The velocity potential may also be used to make a prediction of CMB anisotropy under the assumption that the peculiar velocities are generated by gravity. Bertschinger, Górski and Dekel (1990) show how one may convert the velocity potential to the gravitational potential and then extrapolate this back in time to obtain the gravitational potential fluctuations present when hydrogen recombined and the universe became transparent to the CMB. Potential fluctuations present at this time generate CMB anisotropy via the Sachs-Wolfe effect (Sachs and Wolfe 1967). Bertschinger *et al.* "predict" that a very distant observer, who is viewing the progenitor of the GA as it was at the time of recombination, would see a maximum Sachs-Wolfe anisotropy of $\Delta T/T = 2 \times 10^{-5}$ on an angular scale of 1 degree, if $\Omega = 1$. If $\Omega < 1$, the anisotropy has larger amplitude and is seen on a smaller angular scale. This prediction neglects Doppler and adiabatic contributions to the anisotropy, which are dominant for angular scales much less than a degree and are expected to be similar in magnitude to the Sachs-Wolfe contribution on the 1 degree scale. It will be difficult to push observational anisotropy limits below 1×10^{-5} on the degree scale, but unless the nearby universe is significantly different from the universe at the distance of our present horizon, at least there is now the firm prediction from gravitational instability theory, independently of any particular theoretical models like cold dark matter, that primordial fluctuations should be detected at this level.

7. CONCLUSIONS

It is clear that measurements and analysis of peculiar velocities have developed into a rich industry. Most workers in the field now accept the reality of at least some of the measured large-scale peculiar motions. This author has high hopes that future work will allow us eventually to reach the goals set forth in Section 1. This is a major development in extragalactic astronomy, which certainly could not have been foreseen when Rubin *et al.* (1976) first announced evidence for large-scale motions nor when Burstein *et al.* (1986) first

presented their results.

While there is little doubt of the promise afforded by peculiar velocity studies, we must be careful about concluding too much from the present state of the measurements and their analysis. All peculiar velocity measurements have large random errors and all suffer from systematic errors such as Malmquist bias and zero-point errors, and probably from environmental correlations even though there are now several lines of evidence that environmental effects cannot entirely explain away peculiar velocities. It is very important that observers and theorists alike keep these errors in mind and that they try to perform their observations and analysis in a way that minimizes errors and allows them to be estimated and, even more importantly, allows these error estimates to be checked independently.

8. REFERENCES

Aaronson, M., Bothun, G. D., Cornell, M. E., Dawe, J. A., Dickens, R. J., Hall, P. J., Sheng, H. M., Huchra, J. P., Lucey, J. R., Mould, J. R., Murray, J. D., Schommer, R. A., and Wright, A. E. 1989, *Ap. J.*, **338**, 654.

Aaronson, M., Bothun, G. D., Mould, J. R., Huchra, J. Schommer, R. A., and Cornell, M. E. 1986, *Ap. J.*, **302**, 536.

Aaronson, M., Huchra, J., and Mould, J. 1979, *Ap. J.*, **229**, 1.

Aaronson, M., Huchra, J., Mould, J. R., Tully, R. B., Fisher, J. R., van Woerden, H., Goss, W. M., Chamaraux, P., Mebold, U., Siegman, B., Berriman, G., and Persson, S. E. 1982a, *Ap. J. Supplement*, **50**, 241.

Aaronson, M., Huchra, J., Mould, J., Schechter, P. L., and Tully, R. B. 1982b, *Ap. J.*, **258**, 64.

Bertschinger, E., and Dekel, A. 1989, *Ap. J. (Letters)*, **336**, L5.

Bertschinger, E., Dekel, A., Faber, S. M., Dressler, A., and Burstein, D. 1990, *Ap. J.*, in press.

Bertschinger, E., Górski, K., and Dekel, A. 1990, *Nature,* in press.

Bertschinger, E., and Juszkiewicz, R. 1988, *Ap. J. (Letters)*, **334**, L59.

Bothun, G. D., Aaronson, M., Schommer, R. A., Huchra, J., and Mould, J. R. 1984, *Ap. J.*, **278**, 475.

Bothun, G. D., West, M. J., Mould, J. R., and Schommer, R. A. 1990, *Ap. J.*, **353**, 344.

Bottinelli, L., Gouguenheim, L., Paturel, G., and Teerikorpi, P. 1988, *Ap. J.*, **328**, 4.

Burstein, D. 1989, private communication.

Burstein, D. 1990, *Rep. Prog. Phys.*, in press.

Burstein, D., Davies, R. L., Dressler, A., Faber, S. M., Lynden-Bell, D., Terlevich, R. J., and Wegner, G. 1986, in *Galaxy Distances and Deviations from Universal Expansion,* ed. B. F. Madore and R. B. Tully (Boston: Reidel), pp. 123.

Burstein, D., Faber, S. M., and Dressler, A. 1990, *Ap. J.*, **354**, 18.

Burstein, D., and Raychaudhury, S. 1989, *Ap. J.*, **343**, 18.

Davis, M., and Peebles, P. J. E. 1983, *Ann. Rev. Astron. Ap.*, **21**, 109.

Dekel, A. 1988, in *Large Scale Motions in the Universe,* ed. V. C. Rubin and G. V. Coyne (Princeton: Princeton University Press).

Dekel, A., and Bertschinger, E. 1990, in *Large Scale Structure and Peculiar Motions in the Universe,* ed. D. W. Latham and L. N. da Costa (San Francisco: Astronomical Society of the Pacific), in press.

Dekel, A., Bertschinger, E., and Faber, S. M. 1990, *Ap. J.,* in press.

Dekel, A., and Rees, M. J. 1987, *Nature,* **326**, 455.

Djorgovski, S., and Davis, M. 1987, *Ap. J.,* **313**, 59.

Djorgovski, S., de Carvalho, R. and Han, M.-S. 1989, in *The Extragalactic Distance Scale,* ed. S. van den Bergh and C. J. Pritchet (Provo: Astronomical Society of the Pacific), pp. 329.

Donnelly, R. H., Faber, S. M., and O'Connell, R. M. 1990, *Ap. J.,* **354**, 52.

Dressler, A. 1987, *Sci. Am.,* **225**, 46.

Dressler, A. 1987, *Ap. J.,* **317**, 1.

Dressler, A. 1988, *Ap. J.,* **329**, 519.

Dressler, A., and Faber, S. M. 1990a, *Ap. J.,* **354**, 13.

Dressler, A., and Faber, S. M. 1990b, *Ap. J. (Letters),* **354**, L45.

Dressler, A., Lynden-Bell, D., Burstein, D., Davies, R. L., Faber, S. M., Terlevich, and Wegner, G. 1987a, *Ap. J.,* **313**, 42.

Dressler, A., Faber, S. M., Burstein, D., Davies, R. L., Lynden-Bell, D., Terlevich, and Wegner, G. 1987b, *Ap. J. (Letters),* **313**, L37.

Faber, S. M., and Burstein, D. 1988, in *Large Scale Motions in the Universe,* ed. V. C. Rubin and G. V. Coyne (Princeton: Princeton University Press), p. 115–167.

Faber, S. M., Wegner, G., Burstein, D., Davies, R. L., Dressler, A., Lynden-Bell, D., and Terlevich, R. J. 1989, *Ap. J. Suppl.,* **69**, 763.

Fall, S. M., and Jones, B. J. T. 1976, *Nature,* **262**, 457–460.

Frenk, C. S., White, S. D. M., Davis, M., and Efstathiou, G. 1988, *Ap. J.,* **327**, 507.

Giuricin, G., Mardirossian, F., Mezzetti, M., and Pisani, A. 1989, *Ap. J.,* **345**, 101.

Górski, K., Davis, M., Strauss, M. A., White, S. D. M., and Yahil, A. 1989, *Ap. J.,* **344**, 1.

Groth, E. J., Juszkiewicz , R., and Ostriker, J. P. 1989, *Ap. J.,* **346**, 558.

Gunn, J. E. 1989, in *The Extragalactic Distance Scale,* ed. S. van den Bergh and C. J. Pritchet (Provo: Astronomical Society of the Pacific), pp. 344.

Haynes, M. P., and Giovanelli, R. 1988, in *Large Scale Motions in the Universe,* ed. V. C. Rubin and G. V. Coyne (Princeton: Princeton University Press).

Juszkiewicz, R., Bertschinger, E., Dekel, A., and Yahil, A. 1990, in preparation.

Kaiser, N. 1988, *M.N.R.A.S.,* **231**, 149–167.

Kraan-Korteweg, R. C., Cameron, L. M., and Tammann, G. A. 1988, *Ap. J.*, **331**, 620.

Lightman, A. P., and Schechter, P. L. 1990, *Ap. J. Suppl.*, in press.

Lucey, J. R., and Carter, D. 1988, *M.N.R.A.S.*, **235**, 1177.

Lucey, J. R., Currie, M. J., and Dickens, R. J. 1986, *M.N.R.A.S.*, **222**, 427.

Lynden-Bell, D., Faber, S. M., Burstein, D., Davies, R. L., Dressler, A., Terlevich, R. J., and Wegner, G. 1988, *Ap.J.*, **326**, 19.

Mould, J., Han, M., and Bothun, G. 1989, *Ap. J.*, **347**, 112.

Ostriker, J. P., and Suto, S. 1990, *Ap. J.*, **348**, 378.

Peebles, P. J. E. 1969, *Ap. J.*, **155**, 393.

Pierce, M. J., and Tully, R. B. 1988, *Ap. J.*, **330**, 579.

Raychaudhury, S. 1989, *Nature*, **000**, 000.

Regős, E., and Szalay, A. S. 1989, *Ap. J.*, **345**, 627.

Rubin, V. C. 1989, in *Le Monde des Galaxies, A Symposium Honoring G. and A. de Vaucouleurs,* ed. H. G. Corsiin and L. Bottinelli.

Rubin, V. C., Ford, W. K. Jr., Thonnard, N., Roberts, M. S., and Graham, J. A. 1976, *Astron. J.*, **81**, 687.

Sachs, R. K., and Wolfe, A. M. 1967, *Ap. J.*, **147**, 73.

Scaramella, R., Basiesi-Pillastrini, G., Chincarini, G., Vettolani, G., and Zamorani, G. 1989, *Nature*, **338**, 562.

Schechter, P. L. 1977, *Astron. J.*, **82**, 569.

Schechter, P. L. 1980, *Astron. J.*, **85**, 801.

Shapley, H. 1933, *Harvard Bull.*, **890**, 1.

Silk, J. 1989, *Ap. J. (Letters)*, **345**, L1.

Staveley-Smith, L., and Davies, R. D. 1989, *M.N.R.A.S.*, **241**, 787.

Strauss, M. 1989, Ph. D. thesis, U. C. Berkeley.

Strauss, M., and Davis, M. 1988, in *Large-Scale Motions in the Universe,* ed. V. C. Rubin and G. V. Coyne (Princeton: Princeton University Press), pp. 255.

Strauss, M., and Davis, M. 1990, in *Large Scale Structure and Peculiar Motions in the Universe,* ed. D. W. Latham and L. N. da Costa (San Francisco: Astronomical Society of the Pacific), in press.

Teerikorpi, P. 1987, *Astron. Ap.*, **173**, 39.

Tonry, J. L., Ajhar, E. A., and Luppino, G. A. 1989, *Ap. J. (Letters)*, **346**, L57.

Tully, R. B., and Fisher, J. R. 1977, *Astron. Ap.*, **54**, 661–673.

Tully, R. B., and Fisher, J. R. 1987, *Nearby Galaxies Atlas* (Cambridge University Press: Cambridge).

Willick, J. A. 1990, *Ap. J. (Letters)*, **351**, L5.

Yahil, A. 1988, in *Large-Scale Motions in the Universe,* ed. V. C. Rubin and G. V. Coyne (Princeton: Princeton University Press), pp. 219.

Yahil, A. 1990a, in *Large Scale Structure and Peculiar Motions in the Universe,* ed. D. W. Latham and L. N. da Costa (San Francisco: Astronomical Society of the Pacific), in press.

Yahil, A. 1990b, these proceedings.

TIDAL INTERACTIONS AND GALAXY
SEGREGATION IN CLUSTERS

F. Combes

DEMIRM, Observatoire de Meudon, F-92190 MEUDON, FRANCE
and Ecole Normale Supérieure, 24 Rue Lhomond, F-75005 PARIS

ABSTRACT
Morphology and luminosity segregations of galaxies in clusters would be crucial observational clues for theories of galaxy formation. If the first one is a widely accepted fact, there is not much evidence for the second. To account for morphological segregation, I review (and bring new) arguments in favor of tidal interactions and mergers after the virialisation of the cluster, in opposition to the more frequently invoked explanation involving the intergalactic hot medium. Galaxy interactions are efficient enough to blur any segregation that could exist before cluster formation.

i

In recent years, luminosity segregation of galaxies has been actively searched for to test and justify theoretical ideas about biased galaxy formation. However, only morphological segregation is observed non- ambiguously.

1- Morphological segregation:

Morphological segregation is a widely accepted observational fact. Dressler (1980) has clearly demonstrated that the fraction of all morphological types depended smoothly on local density. He has studied 55 rich clusters of galaxy, and derived the famous curves of fig.1a. The fact that they are plotted in function of the local surface density of galaxies and not of the distance to the cluster-center, greatly improves the correlation: populations are clearly a function of local rather than global conditions.

These results were confirmed and extended by Bhavsar (1981) to smaller groups of galaxies, and by Postman, Geller (1984) to 6 orders of magnitude in space density. Giovanelli et al (1986) again supported them by detailed observations of the Pisces-Perseus clusters.
Recently, Binggeli et al (1990) found the same morphological segregation among dwarfs: gas-rich dwarfs avoid the high-density regions, while dwarf elliptical are more clustered (fig 1b).
In "bottom-up" scenarios, this morphological segregation is possible only by environment dependent mechanisms, during or after virialisation of clusters.

2-Luminosity (or mass?) segregation:

There are several reasons to expect luminosity segregation:
1) Two-body relaxation between galaxies in clusters would lead to equipartition: by dynamical friction, the more massive galaxies should fall towards the center.
2) According to biased galaxy formation, diffuse dwarfs should be seen in the voids (Dekel & Silk 1986). Indeed, they originate from small (1σ) density fluctuations, and their gravitational well is not enough to retain gas after the first starburst. While the brighter galaxies have condensed from higher density perturbations (3σ), and should be expected more frequently in clusters and superclusters.

Although expected by several theories, there is no real evidence for luminosity segregation. The dwarfs distribution has recently been the object of much controversy. Davis and Djorgovski (1985) have claimed that low surface brightness galaxies in the UGC (Uppsala General Catalogue) are less clustered than brighter galaxies. But this result in fact was due to the different depths probed by their dwarf and normal galaxies sample. White et al (1988) arrived to the same conclusions with the NGC galaxies catalogued from an HI survey (Tully, 1988). But, as claimed by Eder et al (1989), this result is flawn by the incompleteness of the sample. Most of the redshifts of the dwarf galaxies come from 21cm data, and dwarfs in clusters are gas deficient.

More numerous are the authors who, like Eder et al (1989), find the same clustering properties for dwarfs and giant galaxies (Bothun et al, 1986; Thuan et al, 1987). Binggeli et al (1990) find also some evidence for the opposite effect: dEs (dwarf ellipticals) even more clustered than bright galaxies. They also remark that dEs are always companions of giant galaxies. The formation of dEs could be intimately related to tides.
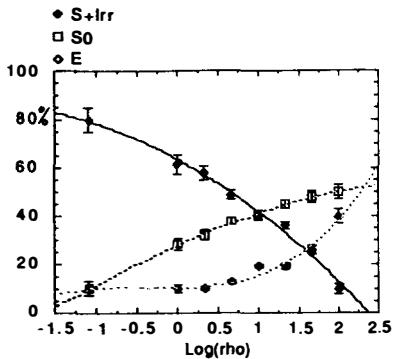
Fig 1a): The fraction of E, S0 and S+Irr galaxies as a function of the log of the projected density, in galaxies Mpc^{-2}. From Dressler (1980).
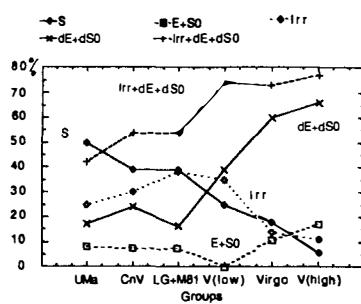
Fig 1b): The fraction of the various morphological types in several poor groups (UMa, CnV), the Local (LG) and M81 groups, and Virgo (low, average, and high density). From Binggeli et al (1990).
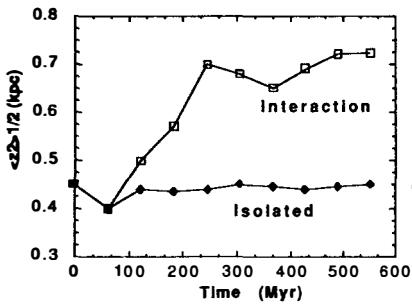

Fig 2: Mean disc thickness averaged over radius, as a function of time, for the control run (isolated galaxy), and for the tidal interaction simulation. From Gerin et al (1990).
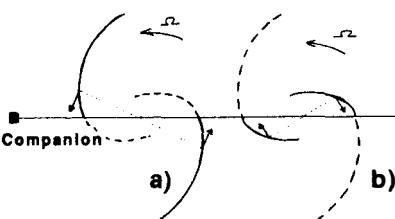
Fig 3: Schematic representation of the directions and signs of the tidal tangential forces applied to the gas in the induced spiral arms. a) outside corotation, the gas will gain angular momentum and will be stripped outwards; b) inside corotation, the gas will flow inwards.

3-Physical Mechanisms

What are the agents responsible for this firmly-stated morphological segregation? Let us stress again that the clear correlation of galaxy types with density is obtained with *local* density only.

The first and most popular explanation invokes the role of the **intergalactic medium** (IGM). Galaxies are moving at high speed in the hot intra-cluster gas, detected by X-ray observations. The ram pressure would then be sufficient to deprive the spiral galaxies from their gas, stopping all new star formation. The spiral galaxies could thus be transformed into lenticulars (Gunn and Gott 1972). Or the interstellar cold medium could evaporate into the hot intra-cluster medium (Cowie and Songaila 1977). These hypotheses however encounter several objections, as summarised by Dressler (1980):
-there is a smooth relation between density and morphological type: lenticulars are found regularly in clusters, even outside the regions filled with IGM. There is no sudden threshold with the minimum IGM density.
-the same morphological segregation holds in groups and un-relaxed clusters, where there is no (or much less) IGM.
-bulges are systematically larger in S0 than in spirals (the bulge-to-disk luminosity ratio should be unchanged after gas ablation by the IGM).

To be compatible with these observations, we must consider a mechanism that could act more gradually, and also both on the stellar and gaseous components. These characteristics are typical of **tidal interactions and mergers.**
Galaxy-galaxy interactions are commonly thought to be inefficient in clusters, due to the high relative velocity of the galaxies. However, the increased density multiplies the number of collisions, such that smaller impact parameters are more frequent than in the field (Combes et al 1988). Already Icke (1985) suggested that repeated tidal stripping of spirals could transform them into S0. Also spheroidal systems could be the result of accumulated mergers: ellipticals as well as S0 bulges.

Let us summary the **effects of an interaction:**
-in only one non-coplanar passage, we can notice already a high thickening of the disk (Gerin et al 1990), as is shown in fig 2.
-tides are equivalent to a $\cos(2\theta)$ perturbation in the plane of the victim-galaxy (fig.3). This bisymmetric perturbation creates a 2-arms spiral structure in the gaseous component; then torques are exerted on the non-axisymmetric structure, the sign of which depends on the relative position of the spiral with respect to the perturbation. This phenomenon is very similar to what happens with a bar perturbation (Combes, 1988). Inside corotation, the gas is driven towards the center.
-outside corotation the gas is stripped outwards and lost in the background (Combes et al 1990).

To demonstrate the efficiency of these tides even for hyperbolic encounter, I have run an interaction **simulation** where the companion is launched in an unbound orbit, whose hyperbolic equivalent would be of excentricity of e=6. The pericenter was about 15 kpc, with a maximum velocity of 700 km/s.
The target-galaxy is represented by both self-gravitating stellar and gaseous components. Their self-gravity is computed by a 2D N-body code, based on the Fast Fourier Transforms method. The FFT are calculated on a 256x256 grid, but the useful space is restricted to 128x128 cells, to avoid images. The cell has size of 0.25 kpc. There are $4 \ 10^4$ particles representing the stellar component, and
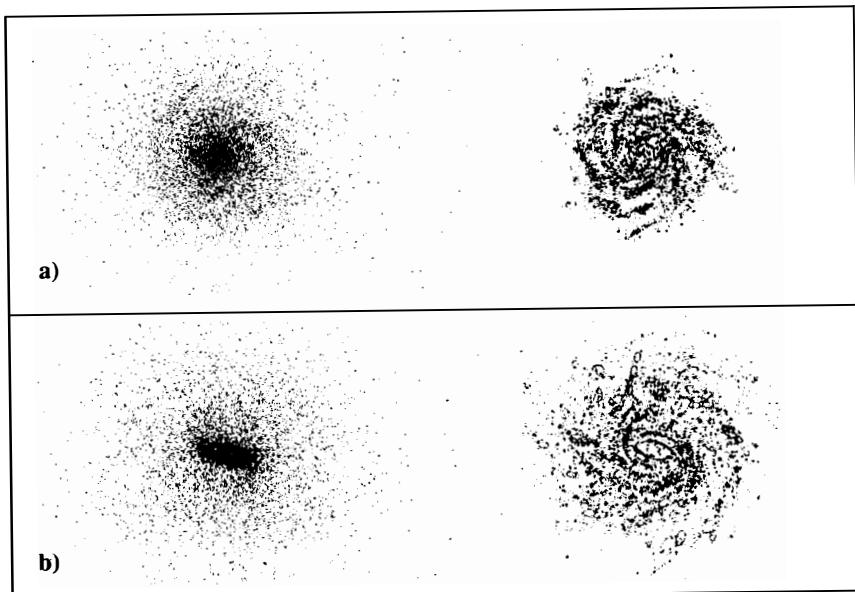
Fig 4: a) 2 10^8 yrs and b) 7 10^9 yrs plots for an isolated galaxy: left=stars, right= gas
c) and d) same for a tidally-perturbed galaxy (companion in an hyperbolic orbit).
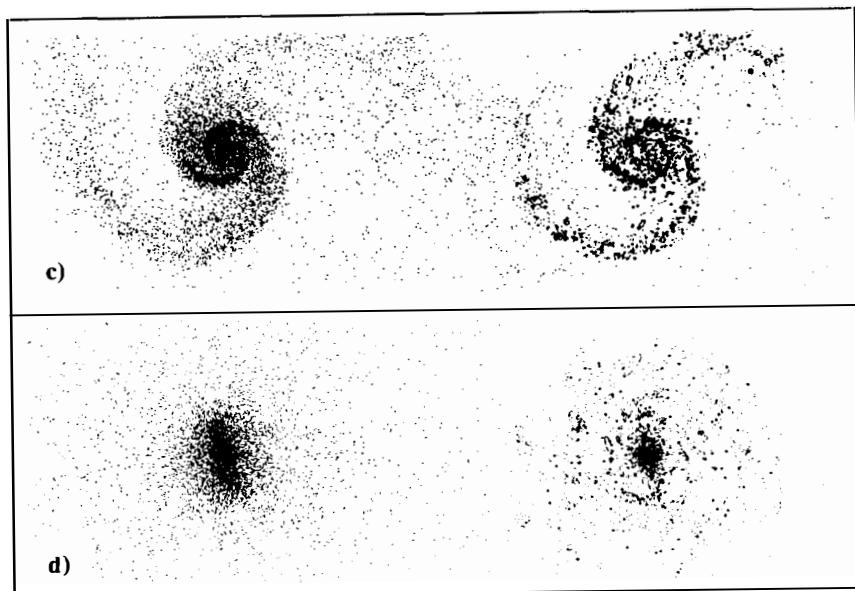
between 10^3 and $2 \cdot 10^3$ gaseous clouds.

The gas clouds ensemble is simulated according to the code developped by Combes & Gerin (1985). The clouds are spread about a mass spectrum, from 10^4 to 10^7 M_0; the latter represents the Giant Molecular Clouds (GMC). Small clouds grow through collisions. The life-time of GMC against dispersion by star-formation is $4 \cdot 10^7$ yrs. Kinetic energy is re-injected in the gas component when a GMC is dispersed. The cell (spatial resolution) of the collision grid is 130pc (240x240). The mass fraction in the gas is $M_{gas}/M_{tot} = 3.6\%$.

Fig. 4 compares the results obtained in the control run (isolated galaxy) with the interaction simulation: the spiral galaxy in the second case is deprived of its gaseous disk, but also part of its stellar disk.

As a side effect, tidal interactions favorise bar instability. We should therefore expect more barred systems in clusters: there is indeed some observational evidence for that (Thompson 1981). Also, if spiral galaxies in clusters are stripped of their HI gas, the molecular gas usually confined inside corotation should not be depleted, but only more centrally- concentrated. In their survey of the Virgo cluster in molecules, Kenney and Young (1988) indeed found normal H_2 global content.

Frequency of interactions: Order-of-magnitude estimates based on geometric cross sections and number density of galaxies show that tidal interactions are sufficient to transform spirals to lenticulars, even with highly hyperbolic encounters. However a more precise estimation requests the knowledge of galaxy orbits since the cluster virialisation, therefore needs a modelisation of the cluster formation.

Icke (1985) has succeeded in reproducing the observed correlations between morphological types and density (fig. 5). The smooth dependence of the correlations (relative fractions vary as density to the 1/6 power) can be retrieved by dimensional arguments:

The number of events able to transform S to S0 are proportional to:
$$dN \ \alpha \ 2 \pi \, b \, db \ n \ \sigma \ (v/\sigma) \ \delta t$$
where b is the impact parameter (independent of the density n), σ the velocity dispersion, v the galaxian rotational velocity, and δt the time to cross a specific volume V (= 1/n).

The factor (v/σ) symbolizes the collision efficiency. δt can be expressed by $1/(\sigma \, n^{1/3})$, while $n \approx \sigma^2 /4\pi GM R_c^2$ or $\sigma \approx n^{1/2}$ (where M is a typical galaxy mass and R_c is characteristic of the cluster size of density n).
Therefore: $dN \approx n/(n^{1/2} \, n^{1/3}) = n^{1/6}$

4-Some Problems

-Dwarfs

Observers have for long noticed the close similarity of structural parameters between all types of dwarfs: dI, dE, BCD. This has led Davis and Phillips (1988) to suggest transformation of dIrr to dEs, via a BCD (Blue Compact Dwarfs) starburst phase. These transformations could be favored by tidal interactions in clusters, and explain the morphological segregation of dwarfs.

This hypothesis is opposite to the Silk et al (1987) scenario: transformation of dE to dIrr by accretion in clusters of previously ejected gas. This predicts dIrr clustered with giants, but not dE, contrary to observations.

In the frame of the tidal interactions scenario, there remain some problems to be solved. Why the proportion of dwarfs should be higher at high density, as suggested by Binggeli et al (1990)?

We can only propose two arguments:
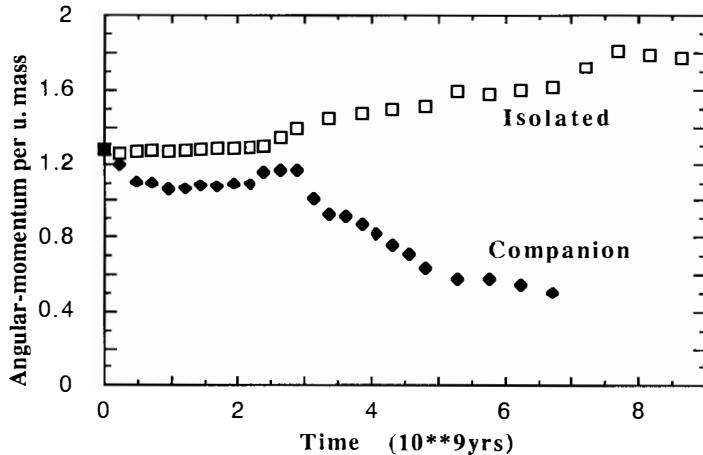
Fig 4: e) Evolution of the angular momentum per unit gas mass for an isolated galaxy and for a tidally-perturbed galaxy (companion in an hyperbolic orbit).
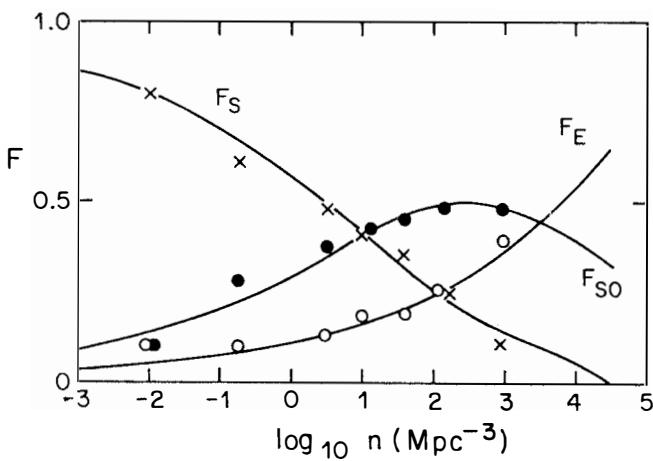

Fig 5: The fraction of all morphological types (S, E, and S0) in function of the projected local density, retrieved from order-of-magnitude estimates by Icke (1985). Curves represent the best model fitted to the observed points from Dressler (1980).

-the merging of spirals giving birth to an elliptical reduces the total number of giant galaxies, automatically raising the percentage of dwarfs.

-mergers could produce dwarfs systems (such as famous examples like the "Antennae").

-Equipartition

Why equipartition expected by galaxy-galaxy interactions is not observed in clusters?

-The clue might be to consider the two concurrent phenomena: equipartition and tidal stripping, which occur on the same time-scales (Merritt 1983-1985). The second stops the first and results in the mass depletion of the core: background particles are expulsed away (to transport the energy). This can lead to mass anti-segregation. But Merritt did not consider the effects of mergers.

-Mergers are a source of bias: larger galaxies are expected towards the densest parts of the cluster (Carlberg 1988, White et al 1987). However,

galaxies of large sizes are excluded from the center (except the very center, cD galaxies) a limitation coming from the cluster tidal field. This may explain the absence of luminosity segregation.

References

Bhavsar S.P. (1981) *Astrophys.J.* **246**, L5

Binggeli B., Tarenghi M., Sandage A. (1990) *Astron. Astrophys.* **228**, 42

Bothun G.D., Beers T.C., Mould J.R., Huchra J.P. (1986) *Astrophys.J.* **308**, 510

Carlberg R.G. (1988) *Astrophys.J.* **332**, 26

Combes F. (1988) in *"Galactic and Extragalactic Star Formation"*, ed. R.E. Pudritz and M. Fich, *NATO ASI series*, Kluwer Academic Publishers, page 475

Combes F., Gerin M. (1985) *Astron. Astrophys.* **150**, 327

Combes F., Dupraz C., Casoli F., Pagani L. (1988) *Astron. Astrophys.* **203**, L9

Combes F., Dupraz C., Gerin M. (1990) Proceedings of Heidelberg Conference (May-June 1989), on *"Dynamics and Interactions of Galaxies"*, ed R. Wielen

Cowie L.L., Songaila A. (1977) *Nature* **266**, 501

Davies J.I., Phillipps S. (1988) *M.N.R.A.S.* **233**, 553

Davis M, Djorgovski S. (1985) *Astrophys.J.* **299**, 15

Dekel A., Silk J. (1986) *Astrophys.J.* **303**, 39

Dressler A. (1980) *Astrophys.J.* **236**, 351

Eder J.A., Schombert J.M., Dekel A., Oemler A. (1989) *Astrophys.J.* **340**, 29

Gerin M., Combes F., Athanassoula E. (1990) *Astron. Astrophys.* in press

Giovanelli A.R., Haynes M.P., Chincarini G.L. (1986) *Astrophys.J.* **300**, 77

Gunn J.E., Gott J.R. (1972) *Astrophys.J.* **176**, 1

Icke V. (1985) *Astron. Astrophys.* **144**, 115

Kenney J. D., Young J. S. (1988) *Astrophys.J. Supp.* **66**, 261

Merritt D. (1983) *Astrophys.J.* **264**, 24; (1984) *Astrophys.J.* **276**, 26; (1985) *Astrophys.J.* **289**, 18

Postman M., Geller M.J. (1984) *Astrophys.J.* **281**, 95

Silk J., Wyse R.F.G., Shields G.A. (1987) *Astrophys.J.* **322**, L59

Thompson L.A., (1981) *Astrophys.J.* **244**, L43

Thuan T.X., Gott J.R., Schneider S.E. (1987) *Astrophys.J.* **315**, L93

Tully R.B. (1988) *Nearby Galaxies Catalog* (Cambridge University Press)

White S.D.M., Davis M., Efstathiou G., Frenk C.S. (1987) *Nature* **330**, 451

White S.D.M., Tully R.B., Davis M. (1988) *Astrophys.J.* **333**, L45

SELF-SIMILARITY
IN
GRAVITATIONNAL COLLAPSE

FABIEN MOUTARDE

Département d'Astrophysique Extragalactique et de Cosmologie,
Observatoire de Paris-Meudon, Meudon, France.

ABSTRACT

As a first step to understand general properties of gravitationnal instability of collisionless matter, we investigate the formation of halos in an $\Omega = 1$ universe. We focus on the non-linear formation of structures at the intersection of caustics. Using numerical simulations we show that a power-law density profile progressively appears, finally reaching the form $\rho \propto r^{-1.8}$ at the collapse time (and $\rho \propto r^{-1.1}$ for 2D collapse). This scale invariance is exhibited before shell-crossing, indicating that it results from the early infall and not from the oscillations of shells through the center. We also derive a new analytical expression giving a better approximation than Zeldovich's, and confirming the simulation results. After the collapse the system soon reaches a self-similar regime, therefore conserving the same power-law profile. The conjectures and predictions of Fillmore and Goldreich[1] on such a regime are tested and mostly verified, except that in our case the slope is simply the one established before collapse. This scale invariance and self-similarity seem to be a generic property of collisionless gravitationnal collapse in an $\Omega = 1$ universe, and might be a clue towards the explanation of global scale invariance observed in the distribution of galaxies.

I Introduction

It is now well established that the major part - more than 90% - of the mass in the Universe consists in dark matter. Moreover, theoretical predictions (inflationnary scenarii) and observationnal estimations (*e.g.* Bertschinger, this volume) of the density parameter Ω, when compared to the primordial nucleosynthesis constraints (*e.g.* Olive, this volume), leads to the assumption that most of this dark matter is non-baryonic. It is also a reasonable hypothesis that the central process responsible for the formation of the large scale structures of the Universe is gravitationnal instability in an expanding background. In this "standard" picture, the non-baryonic matter is a collisionless and dissipationless fluid dominating the gravitationnal field in which ordinary matter falls, eventually forming stars and galaxies. In order to test this scenario one has first to understand the gravitationnal dynamics of such a collisionless fluid. The linear stage can be dealt with analytically, but the non-linear phase can be studied only with approximations or numerical simulations. Most of the numerical works aimed at reproducing a whole part of universe in a given scenario (Hot Dark Matter, Cold Dark Matter) and then compare it to the real Universe (*e.g.* White et al.[2]). This approach allows to discriminate between different types of initial fluctuations on the basis of global properties, but gives little insight on the nonlinear gravitationnal collapse itself. An alternative approach consists in studying the collapse around simple isolated overdensities in order to infer some general properties of the process. This was done for instance by Melott[3] for 1D perturbations (pancakes). We report here on a similar work for the formation of filaments and clusters. In order to separate the various effects at play we first focus on the idealized case where only one scale is present, so that the objects we are studying are forming at the intersection of caustics. Our work is based on high resolution Particle-In-Cell simulations of the non-linear gravitationnal collapse of collisionless matter in an $\Omega = 1$ expanding universe. In this new approach of the halo formation problem we were particularly interested in checking the time and condition of appearance of the self-similar solutions found by Fillmore and Goldreich[1] (hereafter FG) and Bertschinger[4] (hereafter B85).

II Precollapse scale invariance

II.1 Numerical simulations

For the numerical simulations we use the Particle-In-Cell code initially developped by Bouchet et al.[5], optimized and used in Alimi et al.[6], and recently transformed into a tridimensionnal code. The initial conditions are obtained as follows: we first place the particles on a regular lattice, generating a purely uniform background, and then imprint the desired perturbation by displacing the particles according to the "Zeldovich algorithm". We use $128^3 = 2097152$ particles in a 128^3 mesh (or 1024^2 particles in a 512^2 grid for simulations of bidimensionnal collapse). The perturbation we use consists in small amplitude (5%) sine displacement along each axis, producing a smooth and compensated cosine density enhancement at the center of the box. The initial velocities are those of the growing mode. Fore more details on the numerical resolution and initialization, see Alimi et al.[3] and Moutarde et al.[7].

During the non-linear phase, but before the collapse (*i.e.* before any shell has reached the center)
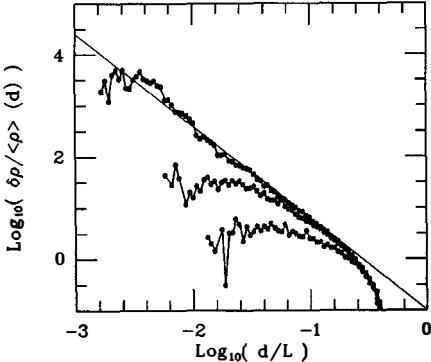
Figure 1: density contrast as a function of comoving radius d (in units of the box length L) for different times during 3D collapse ($a \simeq 0.7a_c^{3D}$, lower curve, $a \simeq 0.9a_c^{3D}$, middle curve, $a \simeq a_c^{3D}$, upper curve). The solid line is the $\rho \propto d^{-1.8}$ fit for the profile at collapse.

a power law density profile progressively appears (see figure 1). At early times the power law part is limited to a small range of scales and its slope is small. As the system evolves the scale invariance propagates towards smaller and smaller scales and the slope increases. The same evolution is seen at the intersection of two caustics. At the collapse time (occuring for an expansion factor $a_c^{3D} \sim 12$ for three caustics[1] and $a_c^{2D} \sim 15$ for two) the density profile is purely scale invariant on more than two decades of scales. It is well fitted by $\rho \propto r^{-1.8}$ for the 3D perturbation and $\rho \propto r^{-1.1}$ for the 2D case. It is capital to note that the scale invariance is obtained *before shell-crossing*, so that it does not result from some asymptotic behavior of the successive shells oscillating through the center (as in the semi-analytical approaches of FG or Bertschinger[4]). On the contrary it is due to the infall of matter during the early formation of the condensation.

II.2 Analytical approximations

As noted in the precedent section the scale-invariance occurs before shell-crossing, so we tried to confirm analytically the simulations results. For non purely spherical collapse, the only analytic description of the non-linear phase is the "Zeldovich approximation" (Zeldovich[8]). It consists in extrapolating the linear theory in the form of trajectories as a function time and Lagrangian coordinate:

$$\vec{r}(t, \vec{q}) = a(t)\vec{q} + b(t)\vec{p}(\vec{q}) \tag{1}$$

where $\vec{q}$ is the Lagrangian coordinate, a(t) is the expansion factor, $\vec{p}(\vec{q})$ is the initial perturbation from uniformity (at $a(t_0) = 1$) and b(t) is the universal time evolution of perturbations in the linear theory (in an $\Omega = 1$ cosmology, $b(t) = [a(t)]^2$ for the growing mode). The condition for equation (1) to be the exact decription of the motion is that the two following equations are satisfied:

$$\vec{\nabla}_r \cdot \left(\frac{\partial^2 \vec{r}}{\partial t^2}\right) = -\Delta_r \Phi = -4\pi G\rho \tag{2}$$

---

[1]the collapse occurs slightly later than $a_c^{Sph} = (3\pi/2)^{2/3} (5/3\delta_*)^{-1} \sim 10$ expected for a purely spherical collapse, because with our initial overdensity, the collapse is spherical only at the first order in the center.

444

$$\vec{\nabla}_r \times \left(\frac{\partial^2 \vec{r}}{\partial t^2} \right) = \vec{0} \tag{3}$$

equation (3) being the condition for the accelerations to derive from a potential. When the perturbation is unidimensionnal, (2) and (3) are fullfilled and equation (1) is the exact solution until trajectories cross each other. The corresponding density profile progressively grows to the $\rho \propto r^{-2/3}$ power-law at collapse near the plane of the pancake (Zeldovich[8]). For 2D or 3D perturbations, equation (1) is only approximate. However, power law density profiles with same slopes as in our numerical simulations are obtained, but not at the right time because the collapse occurs only at $a(t) = 20$ for the "Zeldovich approximation" instead of $a(t) = a_c^{3D} \sim 12$. This motivated us to compute higher order analytical approximations. We used a perturbative method, adding to the right hand side of (1) terms determined by the requirement that (2) and (3) be satisfied at a higher order (with respect to the magnitude of the initial perturbation). We applied this method in our particular case of initial perturbations, and deduced a third order approximation (see Moutarde et al.[7]) for the formula and more details on the calculation).
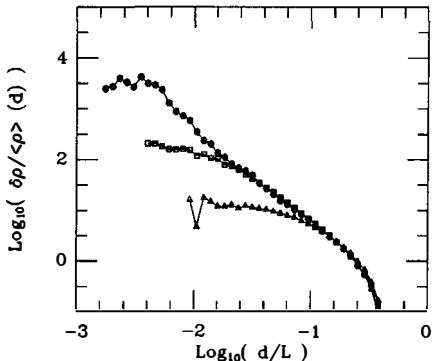
Figure 2: comparison of the density profiles at the same expansion factor $a \simeq a_c^{3D}$ for the Zeldovich approximation (open triangles), our third order approximation (open squares) and for the numerical simulation (filled circles).

In order to compare the various approximations with the simulations we took exactly the same set of particles as in the simulations, moved them according to the analytical formulae, and computed the radial density profile as for the simulation. In figure 2, we compare the simulation, the "Zeldovich approximation" (which is the first order in our perturbative method), and our third order expression at the same expansion factor $a = a_c^{3D} \sim 12$. One can see that our approximation perfectly matches the simulation up to density contrast $\delta \lesssim 50$ and thus exhibits on this domain the same scale-invariant profile. On the other hand the first order expression is much less evolved at this time, diverging from the simulation for $\delta \gtrsim 5$. It should be stressed that the higher order approximations do not consist in a simple correction of the collapse time: we checked that the different approximations are not the same *before the collapse*, even when compared at the same fraction a/a_{coll} of their respective collapse times. Furthermore the exactitude of the collapse time should be crucial for more complicated initial perturbations (*e.g.* a CDM spectrum realization) where different structures of different scales and locations are expected to collapse at different

times.

Nevertheless, the important point for the present study is that the precollapse power-law density profile found with the numerical simulation is confirmed analytically by kinematic type approximations.

III Self-similar regime

As we have mentioned above, the collapse of a pancake does yield a power-law density profile at the collapse time, but afterwards it is replaced by a succession of high density spikes (Melott[3]). On the contrary, the precollapse scale invariance at the intersection of two or three caustics is most interesting since the subsequent evolution quickly becomes self-similar. As shown on figure 3, in the case of 3D collapse,the density profile is then a power-law in space and time $\rho_{com} \propto d^{-1.8} t^{0.8}$ (ρ_{com} being the comoving density and d the comoving radius). The self-similarity can also be seen on the peculiar velocities: the r.m.s. value grows as $t^{0.13}$, and the peculiar velocity distribution is invariant when scaled to this r.m.s. value: the histogram $N(v/ < v^2 >^{1/2})$ does not depend on time during the self-similar regime. For the filament formation (2D collapse), the density scales as $\rho_{com} \propto d^{-1.1} t^{0.37}$, and $< v_{pec}^2 >^{1/2} \propto t^{-0.2}$ (see Alimi et al.[6]) for more details).
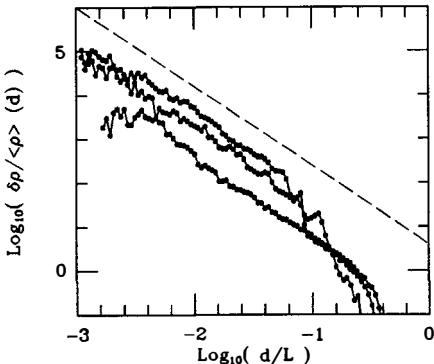
Figure 3: comoving density profiles at different times after the collapse ($a \simeq a_c^{3D}$, lower curve, $a \simeq 1.7 a_c^{3D}$, middle curve, $a \simeq 3.3 a_c^{3D}$, upper curve). A $\rho \propto d^{-1.8}$ dashed line is given for comparison.

Gunn[9] had based his analytic study of secondary infall on the crude hypothesis that after a shell of matter separated from the general expansion (an instant named its "turnaround"), it would oscillate through the center with a maximum radius (denominated "apapsis") fixed to a constant fraction of its turnaround radius. In their more general search for self-similar solutions of gravitationnal collapse, FG assumed that the apapsis of all shells have the same power-law time dependance. Even though the collapses we simulate are not purely cylindrical or spherical (cf. typical trajectory, figure 4), each particle oscillates around the center in a manner similar to the radial infall case (cf. radius evolution, figure 5) so that we can measure successive apapsis and study their time dependance. We show in figure 6 that it is indeed very well fitted by a power-law, as postulated by FG. We also checked that the slope is nearly the same for any test particle. In the

446

3D case, it has to be noted that the apapsis is nearly constant in physical coordinates, as obtained semi-analytically by B85 for "hatbox" initial overdensity.
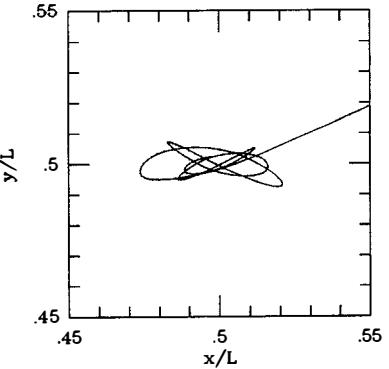
Figure 4: typical trajectory of a particle shown in comoving coordinates (2D collapse).
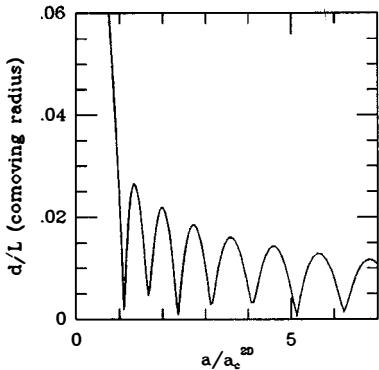
Figure 5: comoving radius d from the center as a function of the expansion factor along the trajectory of figure 4.
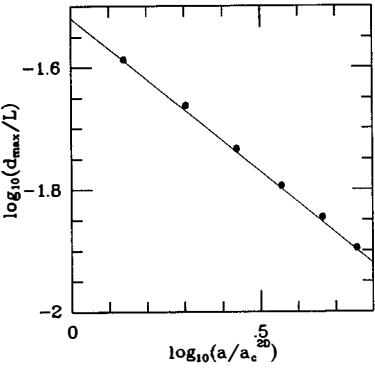
Figure 6: evolution of the apapsis (*i.e.* the successive maximal radii d_{max}) in comoving coordinates. The solid line shows the $d_{max} \propto a^{-0.5}$ fit.

Another conjecture by FG was the existence of an adiabatic invariant $J = \int_0^{r_a} \frac{dr}{dt} dr$ where the integral is taken between two successive extrema of r. Adapting the above expression to our case of non radial trajectories we computed $I = \int_{r_{min}}^{r_{max}} \frac{d\vec{r}}{dt} . d\vec{r}$ for various test particles, and found that it is roughly constant along any given trajectory. Finally, using their various hypothesis FG proved that the parameters of apapsis variation ($r_{max} \propto t^q$) and mass profile ($M \propto r^\gamma t^{-s}$) are related by:

$$s = q(\gamma - n + 4) \tag{4}$$

where n is the dimensionnality (n=2 for cylindical collapse, n=3 for spherical collapse). In our

simulations we measure independantly s, γ and q. At the intersection of two caustics, $q \sim 0.3$, $\gamma \sim 0.9$, which should imply $s = 0.87$ in good agreement with the measured value $s \sim 0.9$. For 3D case, $s \sim q \sim 0$ so that (4) is also well verified.

IV   Discussion

Our work shows that self-similar solutions found and studied by FG and B85 do occur during halo formation. Moreover, we find that this self-similar regime is established very quickly after the collapse, due to the preparation of the system which has a scale invariant density profile just at the collapse. This scale invariance is caused by the early infall and can be reproduced analytically with our new improved Zeldovich-type approximation. The self-similar regime is found to have all the properties predicted or conjectured in the semi-analytic works of FG and B85. In particular there is an adiabatic invariant along the particles trajectories. However, the slope of the density profile is not reached asymptotically, but results from the precollapse phase, keeping some memory of initial conditions.

In order to test if our result is particular to the degenerate case of spherical collapse, we ran several simulations with unequal amplitudes sine perturbations (corresponding to ellipsoidal initial perturbations). The power law is less clean and extended than in the symmetric case, however the qualitative behavior is the same, and the profile established when collapse just occurs in the third direction is well fitted by $\rho \propto r^{-1.8}$ on a large range of scales. We also tested the influence of small scale inhomogeneities or initial velocity dispersion (non zero temperature). In these cases a constant density core is formed in the center, but the rest of the power-law halo remains unaffected (see Alimi et al.[6]). This suggests that the scale invariance and self-similarity properties we have pointed out are general, at least for structures formed at the intersection of caustics. Moreover, we are presently trying different initial overdensities (hatbox, ...) and it seems that the qualitative behavior is also the same: progressive formation of a power-law before collapse, and self-similarity afterwards. But the slope and its validity domain appear to vary with the shape of the initial perturbation. A first consequence is that characteristics of present-day galaxy or cluster halos might give some information on their genitor perturbation. On the other hand, the common qualitative behavior for very different initial conditions suggests that we have pointed out a generic property of collisionless gravitationnal collapse in an $\Omega = 1$ universe.

It is interesting to note that Alimi, Blanchard and Schaeffer[10] have recently shown by a counts-in-cell analysis that the observed distribution of galaxies exhibits a global scale invariance. The same property was also found by Bouchet[11] in Cold Dark Matter simulations. These two works deal with global statistics on regions containing many collapsed structures, while we study properties of isolated halos. However, if local collapses universally obey scale invariance and self-similarity (as suggested by our results), this might be at the origin of the aforementionned global scale invariance.

448

REFERENCES

1) Fillmore, J.A., and Goldreich, P., 1984, *Ap. J.*, **281**, 1 (FG).

2) White, S.D.M., Frenk, C.S., Davis, M., and Efstathiou, G. 1987, *Ap. J.*, **313**, 505.

3) Melott, A.L., 1983, *Ap. J.*, **264**, 59.

4) Bertschinger, E., 1985, *Ap. J. Supp.*, **58**, 39 (B85).

5) Bouchet, F.R., Adam, J.-C., and Pellat, R., 1985, *Astron. A.*, **144**, 413.

6) Alimi, J.-M., Bouchet, F. R., Pellat, R., Sygnet, J.-F., and Moutarde, F., 1990, *Ap. J.*, in press.

7) Moutarde, F., Alimi, J.-M., Bouchet, F. R., Pellat, R., and Ramani, A., 1990, in preparation

8) Zeldovich, Ya. B. 1970, *Astron. A.*, **5**, 84.

9) Gunn, J.E., 1977, *Ap. J.*, **218**, 592.

10) Alimi, J.-M., Blanchard, A., and Schaeffer, R., 1990, *Ap. J. Lett.*, **349**, L5.

11) Bouchet, F.R., 1988, Proceedings of the 8th Moriond Astrophysics Meeting "Dark Matter".

THE ORIENTATION OF GALAXIES IN THE SEPARATE GALAXY

STRUCTURES IN THE LOCAL SUPERCLUSTER

Wlodzimierz Godlowski

Jagiellonian University Observatory

ul.Orla 171 30-244 Kraków Poland

Summary:
     The method of investigating the orientation of galaxy
rotation axes in the Local Supercluster introduced by Jaaniste
and Saar (1977), which takes into account both galaxy position
angle and inclination, has been applied to a sample of 2227
galaxies. I analyze the orientation of galaxies in the whole
Local Supercluster and in the special groups of galaxies. I
found that the distribution of galaxy planes is anisotropic.
Galaxy planes rather tend to be perpendicular to the Local
Supercluster plane. The projection of galaxy rotation axes on
the Local Supercluster plane rather tend to point toward the
Virgo Cluster center. The distributions the spirals galaxies
are different then non-spirals. The results are compared also
with results previous workers.

1.Introduction

The investigation of the distribution of galaxy orientation in the Local Supercluster has a long history. Most previous papers showed that distribution is not uniform. Nearly all from this results could be interpreted as parallelism of galaxy planes to the supergalactic plane. Jaaniste and Saar (1977;JS) found that galaxy rotation axes tend to be distributed parallel to the LSC plane. They analyzed not only the distribution of galaxy position angles ρ, but also took into account their inclination i to the observer's line-of sight. When we studies the distribution of galaxy position angles we may take into account only galaxies for which this parameter has been determined. It mean that all galaxies seen "face-on" or nearly "face-on" should be excluded. However, even for such galaxies important information can be extracted.

The JS is full of inconsistencies as was showed by the preliminary analysis of it. Some of them have already been mentioned in our previous work. Moreover, JS used only angle δ_{D}, giving the orientation of galaxy planes with respect to the LSC plane. We used also the second important parameter, angle η, which is the angle between the projection of rotation axis on the supergalactic plane and the direction toward Virgo Cluster center. We based on the projection of galaxies on the celestial sphere. All analogous analysis of galaxy orientation gives a 4-solution ambiguity. Consideration of rotation axes only, instead of angular momenta, decreases this to two solutions without loosing information about the distribution of galaxy planes with respect to the LSC plane. Using JS approach we may include to our analyze galaxies seen "face-on". It is important because is no a priori reason for discounting those galaxies from the analysis.

2.Observational data.

First of all I make a cross correlation between "A Catalogue of Galaxy Red shifts" (Rood 1980) and catalogues "Upsala General Catalogue of Galaxies" (Nilson 1973, UGC) and "Upsala Survey of ESO" (Lauberts 1982, ESO). Each galaxies with radial velocities (correct for solar motion) below 2600 km/s constitute that sample search in literature for radial

velocities of galaxies. For each galaxy I take from UGC or ESO position on the celestial sphere in equatorial coordinates: α, δ, its magnitude, the diameters a and b of the major and minor axes and the position angle ρ_*. The inclination angle, i.e. the angle between the normal to the galaxy plane and the observer's line-of-sight was computed from the formula: $\cos^2 i = (q^2 - q_o^2)/(1 - q_o^2)$, which is valid for oblate spheroids (Holmberg 1946). We also take into account galaxies denoted in catalogs as face-on. For these galaxies $i = 0$. The values of observed axial ratio $q = b/a$ is based on UGC or ESO data, while the true axial ratios q_o are taken from H^2V (Heidman et al. 1971). For peculiar galaxies it was assumed that $q_o = 0.2$. The sample was divided into three groups according to morphological type; groups containing spirals and lenticular galaxies are denoted as "spirals", the remaining ones are denoted as "non-spirals". Both groups together are denoted as "all". Investigating the distribution of galaxy in the LSC it is convenient to express the galaxy position and theirs position angle in the supergalactic coordinate system, instead of the equatorial system, as in UGC. The supergalactic coordinate system is defined (Flin, Godlowski 1986):

a: the coordinates of the supergalactic pole in the equatorial system are: $\alpha = 285°.5$, $\delta = +16°$

b: the basic great circle "meridian" of the supergalactic system passes through the Virgo Cluster center with coordinates (Sandage and Tammann 1976): $\alpha = 186°.25$, $\delta = +13.1$. In this system we obtain /where $p = \rho - \pi/2$/ :

$\sin\delta_D = -\cos i * \sin B \stackrel{+}{-} \sin i * \cos p * \cos B$

$\sin\eta = (\cos\delta_D)^{-1} * (-\cos i * \cos B * \sin L + \sin i * (\stackrel{-}{+}\cos p * \sin B * \sin L \stackrel{+}{-} \sin p * \cos L)$

3. Angular distribution isotropy tests.

We should check if the distributions of the angles δ_D and η are isotropic. I applied three different statistical tests, following Hawley and Peebles (1975) and Kindez (1986). These are: χ^2 test, the Fourier test and the auto correlation test. In all tests the range of the angle θ (where $\theta = \delta_D + \Pi/2$ or η) was divided into n bins (n=36).

Let N denotes the total number galaxies in the sample, N_K

the number of galaxies in k-th bin N_O the mean number of galaxies per bin and N_{OK} expected number of galaxies in k-th bin. In the χ^2-test the measure of deviation of the observed distribution from the theoretical, isotropic distribution is given by the statistics: $\chi^2 = \sum_1^n (N_K - N*p_K)^2 / N*p_K$ Where p_K is the probability that a chosen galaxy falls into k-th bin.

The Fourier test checks how the departure from isotropy slowly varies with the angle θ. The model is:

$$N(\theta_K) = N_O *(1 + \Delta_{11}*\cos 2\theta_K + \Delta_{21}*\sin 2\theta_K + \Delta_{12}*\cos 4\theta_K + \Delta_{22}*\sin 4\theta_K)$$

$$\Delta_{1J} = \sum_1^n (N_K - N_{OK})*\cos 2J\theta_K / \sum N_{1OK}*\cos^2 2J\theta_K \quad \text{where} \quad \sum N_{1OK}*\cos^2 2J\theta_K \approx \frac{N_O}{2}$$

and $\Delta_{2J} = \sum_1^n (N_K - N_{OK})*\sin 2J\theta_K / \sum_1^n N_{OK}*\sin^2 2J\theta_K$.

The probability that the amplitude $\Delta_J = (\Delta_{1J}^2 + \Delta_{2J}^2)^{1/2}$ is greater than a certain chosen values given by formula $P(>\Delta_J) = \exp(-\frac{n}{4}*N_O \Delta_J^2)$. The standard deviation of the amplitude is $\sigma(\Delta) = (2/n*N_O)^{1/2}$. We may found the direction of departure from isotropy or from Δ_{11} or from coefficient F considering the model $N(\theta_K) = N_O (1 + F*\cos 2\theta_K)$, (where the F-coefficient is obtained from the fit). If $F<0$, $\Delta_{11}<0$ then an excess of galaxies with rotation axes parallel to the supergalactic plane is observed, whereas for $F>0$, $\Delta_{11}>0$ the rotation axes tend to be perpendicular to the plane. The standard deviation of the F-coefficient is: $\sigma_F = (\sum_1^n (F_K - F)^2 / n(n-1))^{1/2}$.

The auto correlation test measures the correlations between the number of galaxies in adjoining angle bins. The correlation function is: $C = \sum_1^n (N_K - N_{OK})(N_{K+1} - N_{OK+1}) / (N_{OK} N_{OK+1})^{1/2}$

In cease isotropic distribution we expect $C=0$. The standard deviation is: $\sigma(C) = n^{1/2}$. The W-coefficient is: $W_B^* = (N_\parallel - N_\perp)/N$ where $N_\parallel$ and $N_\perp$ denote the number of galaxies with rotation axes parallel and perpendicular to the supergalactic plane. $W_B = W_B^* - \overline{W}_B$ where $\overline{W}_B = 2p_\parallel - 1$. The $W_A = (\sum_\parallel N_i / p_i n - \sum_\perp N_i / p_i n)/N$. where $\sigma(W_B) = (4p_\parallel (1-p_\parallel)/N)^{1/2}$ and $\sigma(W_A) \approx 1.56 \sigma(W_B)$.

4. The orientation of rotation axis of galaxies in the Local Supercluster.

4.1 The distribution of the angle δ_D.

Results are presented in the Table 1. The anisotropy is

observed for all and non-spirals galaxies. For "spirals" galaxies situation is not so clear. The shapes of all distributions are of the same character. The signs of the F-coefficient and Δ_{11} are negative (W are positive). It mean that galaxy rotation axes are parallel to the supergalactic plane.

Tab 2 constitute a convincing evidence that distribution of galaxy planes for "face-on" galaxies and for "edge-on" galaxies are not the same. From the structure of the LSC and the observed positions of galaxies it follows that large majority of galaxies is located at low supergalactic latitude. When we take into account only galaxies seen "edge-on", with $|B|<30^o$ we found excess galaxies with rotation axes perpendicular to the LSC plane. The effect is growing when q is falling off. In this sample we excluded some galaxies - mostly with low δ_D. So theoretical isotropic distribution we obtain from computer simulations. Our result supports Jaaniste and Saar's claim concerning the important role of "face-on" galaxies in studies of distribution of galaxy rotation axes.

Tab.1 Distribution angles δ η for the whole LSC

	N	$P(\chi^2)$	$P(\Delta)$	C	Δ_{11}	$\sigma(\Delta)$	W_A	W_B	$\sigma(W_B)$	F	$\sigma(F)$
δ											
ALL	2227	106.	0.00	39.4	-.09	0.02	0.09	0.06	0.01	-.23	0.07
SP	1565	46.	0.43	6.7	-.03	0.03	0.03	0.02	0.02	-.09	0.05
NSP	662	115.	0.00	60.4	-.23	0.04	0.21	0.16	0.03	-.56	0.18
η											
ALL	2227	75.	0.00	37.2	0.12	0.02	-.07	-.07	0.01	0.00	0.08
SP	1565	42.	0.01	5.5	0.08	0.03	-.04	-.04	0.02	-.06	0.08
NSP	662	80.	0.00	38.3	0.24	0.04	-.14	-.14	0.03	0.15	0.11

Tab.2 Distribution angles δ η for edge-on galaxies and $|B|<30^o$

	N	$P(\chi^2)$	$P(\Delta)$	C	Δ_{11}	$\sigma(\Delta)$	W_A	W_B	$\sigma(W_B)$	F	$\sigma(F)$
δ											
q<.75	2114	65.	0.00	23.8	0.08	0.03	-.02	-.04	0.02	0.01	0.09
q<.60	1650	59.	0.00	18.9	0.10	0.04	-.05	-.05	0.02	0.04	0.09
q<.45	1032	68.	0.00	23.3	0.17	0.04	-.10	-.09	0.03	0.05	0.20
q<.30	530	61.	0.00	21.6	0.23	0.06	-.17	-.14	0.04	0.28	0.24
η											
q<.75	2114	43.	0.46	1.2	0.04	0.03	-.01	-.01	0.02	-.04	0.08
q<.60	1650	43.	0.63	-4.2	0.03	0.04	-.01	-.01	0.02	0.02	0.12
q<.45	1032	46.	0.59	-8.6	-.04	0.04	0.03	0.03	0.03	-.06	0.14
q<.30	530	56.	0.17	-3.8	-.10	0.06	0.09	0.08	0.04	-.22	0.23

4.2 The distribution of the angle η.

In the case angle η the anisotropy is observed too. Even

for spirals galaxies we observed weak anisotropy. The sign of the Δ_{11} coefficient are positive, which means that the projection of rotation axes on the supergalactic plane tend to point toward center of the Virgo Cluster. For edge-on galaxies we observed weak anisotropy only for galaxies with q<0.3

4.3 The distribution of the angles δ_D and η for separate galaxy structures.

I has also analyzed the distribution of the angles δ_D and η for separate galaxy structures within the LSC. Group A is the Virgo Cluster itself. Group B is the disc of the local Supercluster, defined as -450; $+650$ km/sec from the supergalactic plane. Group C are the non-disc galaxies. Groups D-I are defined according Tully 1986:

D	E	F
V>1500 km/s	V>1100 km/s	V>1500 km/s
B∈(-75°,-45°)	B∈(-40°,-50°)	B∈(+12°,+14°)
L∈(-20°,+75°)	L∈(-70°,+70°)	L∈(-50°,+60°)

G	H roughly	I
V>750 km/s	V>450 km/s	V>1300 km/s
B∈(+20°,+50°)	B∈(-85°,-20°)	B∈(-85°,-60°)
L∈(+105°,+255°)	L∈(+90°,+270°)	L∈(+90°,+270°)

Tab3 Distributions angles δ η in the separate structures LSC

	N	$P(\chi^2)$	$P(\Delta)$	C	Δ_{11}	$\sigma(\Delta)$	W_A	W_B	$\sigma(W_A)$	F	$\sigma(F)$
δ											
A	1310	124.	0.00	43.1	-.11	0.03	0.09	0.64	0.02	-.20	0.11
B	1341	186.	0.00	109.5	-.18	0.03	0.13	0.08	0.02	-.12	0.09
C	885	89.	0.30	17.4	0.05	0.03	0.02	0.03	0.02	-.40	0.20
D	31	58.	0.03	15.1	0.40	0.18	-.73	-.41	0.12	2.15	0.82
E	201	141.	0.00	52.3	-.07	0.07	-.01	0.12	0.05	-.37	0.27
F	154	60.	0.02	29.7	0.03	0.08	0.02	0.03	0.05	-.25	0.19
G	67	122.	0.00	8.6	-.17	0.12	0.15	0.12	0.08	0.23	0.67
H	37	43.	0.12	-2.8	0.08	0.16	-.19	-.06	0.11	0.09	0.48
I	22	120.	0.01	-12.0	0.48	0.21	-.25	-.19	0.14	0.21	0.73
η											
A	1310	113.	0.00	62.9	0.21	0.03	-.13	-.13	0.02	0.16	0.10
B	1341	60.	0.00	22.9	0.12	0.03	-.06	-.06	0.02	-.01	0.08
C	885	54.	0.00	18.4	0.13	0.03	-.09	-.09	0.02	0.02	0.11
D	31	67.	0.00	17.0	0.40	0.18	-.23	-.23	0.13	0.10	0.55
E	201	65.	0.02	4.7	0.20	0.07	-.17	-.17	0.05	0.19	0.22
F	154	49.	0.17	7.1	0.15	0.08	-.11	-.11	0.06	0.30	0.35
G	67	93.	0.42	4.6	0.14	0.12	-.19	-.19	0.09	0.75	0.53
H	37	52.	0.79	-5.9	0.11	0.16	-.03	-.03	0.12	-.72	0.62
I	22	31.	0.67	7.5	-.03	0.21	-.09	-.09	0.15	0.83	0.54

All result are showed in table 3.

In group D-I anisotropy is rather observed but directions of the departure from isotropy are different in various groups. So the generally observed alignment of galaxy rotation axes with the Supercluster plane results from different structure contributions. Anisotropy is greatest for the disc galaxies then for the non-disc galaxies. Moreover, in the case of η angle analyzes groups positions and values Δ_{11}-coefficient, permits us to say that projection of galaxy rotation axes on the supergalactic plane rather tends to point toward the Virgo Cluster center. The anisotropy is very strong for Virgo Cluster (A). The perpendicularity of galaxy planes to the radius vector to the Virgo Cluster center has been noticed by MacGillivray and Dodd (1984) in the case of the Virgo Cluster itself. Our Tests show that for Virgo Cluster region the values of the Δ_{11}-coefficient are much greater then $\sigma(\Delta)$. This effect is greatest for the Virgo Cluster itself.

5. Discussion.

In our sample there are 2227 galaxies. The total number of galaxies in which any wrong information occurred during checking was below 1%. We don't know of course, the number of errors which remains in basic catalogues, but it was estimated to be 1%. The comparison of different system of diameter measurement in the main galaxy catalogues, carried out by Fisher and Tully (1981) shows that there is good correspondence among them. The observed axial ratios b/a should be converted to the standard, photometric axial ratios which removes the Holmberg effect. For this purpose the relevant formula from Fouque and Paturel (1985) was applied. This results are presented in tables 2 to 5. The sample was also analyzed without Holmberg effect and with galaxies treated as a flat discs (q=0). There was used also other formulas for Holmberg effect. The differences between these cases were statistically negligible. In order to study the influence that accuracy of measurement might have on this results, random variations of the axes and position angles was introduced. Moreover for "face-on" and nearly "face-on" galaxies axial ratios a/b was created randomly. The results obtained in these cases were

similar to the ones with original data. More detailed we discussed our sample in FG 1986.

6. Conclusions

It has been shown that the Jaaniste and Saar approach of taking into account both parameters position angles and the inclination angles, is very adequate for the study galaxy orientation. This approach, moreover, avoids some of the over simplification inherent in the classical one. This method is applied to the sample of 2227 galaxies from the LSC. This sample is uncontaminated by background objects. The distribution of galaxy planes tend to be perpendicular to the LSC plane. The projection of rotation axes on the supergalactic plane tend to point toward center of the Virgo Cluster. The anisotropy is very strong for the Virgo Cluster. This effects are greatest for "non-spirals" then for "spirals galaxies". It should be remembered that detected perpendicularity of galaxy planes to the LSC plane can also be caused by the perpendicularity of galaxy planes to the vector to the LSC center. The preferred orientation depends strongly on the on the analyzed sample of galaxies. The distribution of "face-on" and "edge-on" galaxies is different. The separate galaxy structures in the LSC was analyzed too. The anisotropy is observed but direction of the departure from isotropy are different in various groups.

The dependence of orientation on the morphological type is a curious phenomenon, which we are not able to explain. Within the framework of the tree main scenarios for origin of galaxies (primeval turbulence, hierarchical clustering and pancake model) our result (perpendicularity of galaxy planes to the LSC plane) excludes primeval turbulence and supports the pancake scenario. However it cannot be excluded that complicated structure of the LSC is due to a hybrid model or and to evolutionary effects.

It seems to me that the solution to the puzzle must be based on further analysis of the distribution and properties of galaxies in the LSC. Further investigation of the possible correlation between supergalactic coordinates, galaxy morphological type and preferred orientation is crucial for the real understanding of the LSC properties.

References
de Vaucouleurs,G.,de Vaucouleurs,A,.Corwin,H.G., 1976 Second
    Reference Catalogue of Bright Galaxies, Texas University
    Press, Austin
Fisher,J.R.,Tully,R.B., 1981 Astrophys. Jour Suppl.Ser. 47,139
Flin,P.,Godlowski,W., 1984 in Cluster and Groups of Galaxies
    (eds F.Mardirossian et.al.) p.65, D.Reidel, Dordrecht
Flin,P.,Godlowski,W.,1986 The orientation of galaxies in the
    Local Supercluster Mon.Not.R.Astr.Soc V222.p525
Flin,P.,Godlowski,W.,1989 The Distribution of galaxy planes in
    the Local Supercluster Pisma w Astron. Zurnał. 15. 10 p867
Fouque,P.,Paturel,G., 1985 Astron. Astrophys. 150,192
Hawley,D.L.,Peebles,P.J.E., 1975 Astr.Jour. 80,477
Heidmann,J.,Heidmann,N.,deVaucoulers,G.,1971,Mem.R.Astr.
    Soc.75, 85
Jaaniste,J.,Saar,E., 1977,Tartu Obser.Prreprint A-2
Jaaniste,J.,Saar,E., 1978, in The large scale structures of
    the Universe (eds.M.S.Longair and J.Einasto) p.488
    D.Reidel, Dordrecht (IAU Symp.79)
Kapranidis,S.,Sullivan III,W.T., 1983,Astron. Astroph. 118,33.
Lauberts 1982 ESO/Upsala Survey of the ESO B Galaxies Garching
    b.Munchen. ESO 1982
Nilson,P.1973 Uppsala General Catalogue of Galaxies (Uppsala
    Astr.Obs.Ann.V, vol. 1)
Nilson,P. 1974, Uppsala Astr.Observatory Report No 3
Rood,H.J.,1980 A Catalogue of Galaxy Redshifts (unpublished)
Tully,R.B., 1986 Astrophys. Jour V.303 p25

LARGE SCALE DISTRIBUTION OF DWARF AND LOW-SURFACE-BRIGHTNESS GALAXIES

Jean-Michel Alimi[1]
Trinh X. Thuan[2]

[1] Laboratoire d'Astrophysique Extragalactique et de Cosmologie, Observatoire de Paris, Section de Meudon, F–92195 Meudon Cédex, France.

[2] Astronomy Department, University of Virginia, P.O. Box 3818, University Station, Charlottesville, VA 22903.

Abstract : We have analyzed a completed sample of dwarf-LSB galaxies selected from UGC catalogue. We distinguish the Dwarf-LSB galaxy distribution in redshift space and in real space. Dwarf-LSB galaxies appear slightly less clustered than bright galaxies. Our results rule out certain classes of biased galaxy formation theories which predict an uniform spatial distribution of dwarf-LSB galaxies.

I. INTRODUCTION

It was suggested by some theories of biased galaxy formation[1] that low surface brightness galaxies might cluster as the mass does, and fill in the voids seen in the bright galaxy distribution. Whether or not the observations are in aggreement with this predictions is not yet clear[2,3,4]. To compare statistically the spatial distribution of the high and low surface brightness galaxies and adress such questions, it was essential to have a reasonably complete sample of dwarf and low-surface-brightness (LSB) galaxies, with well-understood selection effects, and with a sky coverage and a depth comparable to that of the CfA survey of bright galaxies. Such sample was assembled from the Uppsala General Catalog of Galaxies[5] by Thuan and Schneider. The aim of this paper is to compare the spatial distribution of dwarf and LSB galaxies to that of bright galaxies. For comparaison with our dwarf and LSB sample, we shall use the magnitude-limited ($m_B \leq 14.5$) bright galaxy sample of the CfA survey.

First, in §II, we discuss difficulties to measure the correlation function of galaxy from a galaxy catalog. In §III we briefly present some theories of biased galaxy formation which predict explicitely the spatial dwarf and LSB galaxy distribution. Finally in §IV we present and discuss several results concerning spatial dwarf and LSB galaxy distribution. We distinguish the analysis in redshift space and in real space.

II. GALAXY DISTRIBUTION AND CORRELATION FUNCTION

From maps of galaxy catalogs, the galaxy distribution clearly appears inhomogeneous. It was proposed many statistical description, for instance the percolation technique[7], but the interpretation of results of such analysis, can be misleading due to the density dependence. White[8] (1979) proposed to use the void probability function. The main emphasis of the probability $P_0(V)$ (of finding a void in a randomly placed sphere of volume V) has been on a scaling law derived from the assumption of hierarchical clustering. An even more powerful tool is the distribution of the counts in cells[9,10], that in principle carries the complete information on the clustering of galaxies. Fractal approach, filling factor... can also be considered. But in definitive, none of these tools met the success of the correlation function. We will discuss in the following the difficulties to estimate the spatial two-point correlation function from a galaxy catalog.

The probabilistic definition of the correlation function[11] is related to the probability δP of finding a neighboring galaxy in a volume δV at a distance r of a random galaxy in the Universe, $\delta P(r) = (1. + \xi_g(r))\bar{n}\delta V$, where $\bar{n}$, the mean number density of galaxies in the universe, is independent of position. The ambiguity does not concern the definition of the two-point correlation function, but the measure of this quantity from a galaxy catalog. In a volume-limited sample, an estimation of ξ_g is given by[12-14], $\xi(r) = \frac{N_{dd}}{N_{dr}} \frac{n_p}{\bar{n}} - 1.$, where N_{dd} is the number of pairs in the samples, N_{dr} the number of pairs between the sample and a random distribution of same geometry with number density n_p. A first problem is the sampling effect, it originates from the smallness of the present catalogues. When the correlation function is estimated from volume-limited samples, strong variations are found within different volumes[14-16]. This reflects the fact that the available surveys are not deep enough to cancel spatial variations. Another important limitation comes from the normalization, that is the estimation of the

mean number density of galaxy in the universe $\bar{n}$, which is generally not equal to the number of particules in the sample divided by the total volume. $\bar{n}$ is estimated from the luminosity function[17]. It would be misleading to use the observed number of galaxies in the sample[14]. Finally the correlation function is perturbed by the peculiar velocity of each galaxy[18]. The correlation length estimated in redshift space is generally larger than the real correlation length. This effect can be partly removed by the cylindrical projection[12,13]. The analysises from both two-dimensional and three-dimensional data[12,19] yield to the remarquable power law behavior of two-point correlation function of bright galaxies, $\xi_g(r) = \left(\frac{r}{r_0}\right)^{-\gamma}$ with $r_0 = 5.4 \ h^{-1}Mpc$ and $\gamma \sim 1.77$. This is believed to hold in the range $0.01 - 10h^{-1}Mpc$.

To study the clustering of galaxies, it is also interesting to consider the ratio α of the cross–correlation functions, especially if one is interested by a segregation effect. In the following we are interested by α_{BD}, the ratio of the J_3 integrals of the cross-correlations between bright galaxies and dwarf-LSB galaxies, that is, the ratio of the mean excess number of dwarf-LSB galaxies around bright galaxies to the mean excess number of bright galaxies around bright galaxies. In the same way we define its counterpart α_{DB}. These quantities[20] are weakly sensitive to sampling fluctuations and to normalisation uncertainties in the measure of the correlation function and in practice weakly sensitive to peculiar velocity effect. In addition, these quantities are easily related to the bias parameter which caracterized the distribution of a given galaxy population (see below).

III. THEORY OF BIASED GALAXY FORMATION

Many properties of the observed galaxy distribution are well reproduced by the Cold Dark Matter (CDM) theory of galaxy formation[21]. In this theory, the power spectrum of the density fluctuations in the early Universe seems to be precisely what is required to explain both the small and the large-scale structure of this distribution. However, to reconcile the value of the cosmic density parameter $\Omega = 1$ predicted by inflation with observations, the galaxies must not trace the mass, and in this sense the distribution of galaxies is said to be biased. In the numerical simulations, for instance, the matter correlation function steepens in time, and the stage of the simulation to be regarded as the present epoch is essentially determined by matching its logarithmic slope to the observed $\gamma \sim 1.77$ of galaxies. The CDM correlation length at this time turns out to be only $r_0 \sim 1(\Omega h^2)^{-1}$. Hence, for the galaxies to match the observed $r_0 = 5.4h^{-1} \ Mpc$, with $\Omega = 1.$, the galaxies must be biased that is the correlation function of bright galaxies ξ_{gg} is amplified with respect to the correlation function of the density field, $\xi_{gg}(r) = b^2 \ \xi_{\rho\rho}(r)$ where b is the bias parameter.

The bias may be due to physical processes that took place at the time of galaxy formation[22,23], and is also a natural feature that arises in hierarchical theories[24-26]. For bright galaxies (i.e., galaxies typically brighter than the magnitude $M = -18.5 + 5\log h$, where h is the Hubble constant in $100 \ \mathrm{km \ s^{-1} \ Mpc^{-1}}$ units) this bias b_B depends on the specific model[27,28] but must lie in the range 1.7 to 2.5 in order to make the dynamical measures of the density parameter compatible with the value of unity predicted by inflationary theories. By contrast, dwarf-LSB galaxies are expected to be a more reliable tracer of the mass distribution[1] than bright galaxies, and their bias parameter b_D should

therefore be much smaller than b_B and closer to unity. Dekel and Silk[1] (1986) argue that the bright galaxies must originate from high density peaks ($2\sigma - 3\sigma$) in the initial fluctuation field, while typical ($\sim 1\sigma$) peaks either cannot make a luminous galaxy at all because the gas is too hot and too dilute to cool in time, or, if only their virial velocity is less than ~ 100 km s^{-1}, they make diffuse dwarf-LSB galaxies by losing a substancial fraction of their mass in supernova-driven winds out of the first burst of star formation. This leads to a selective bias, in which the bright galaxies are biased towards the clusters and superclusters, while the dwarf-LSB galaxies do trace the mass, and should provide an observational clue for the real distribution of the DM.

Davis and Djorgovski[29] (1985) have studied the two-dimensional low surface brightness (LSB) and high surface brightness (HSB) galaxy distribution. The LSB galaxies are distributed more smoothly than are these of HSB. This phenomenon is attributed to biasing. In the opposite Bothum[30] et al. (1986) show the two classes of galaxies have qualitatively the same distribution. But, White[24] et al (1987) have demonstrated qualitative results are misleading. In their numerical simulations for a biased CDM models, dwarfs and giants delineate the large scale structure equally well. Only by applying quantitative measures such as correlation function can the effects of biasing be seen. Many others related works[4,31−33] are for or against biasing. Only a quantitative, tri-dimensional and with a statistic comparable to CfA survey, analysis of dwarf-LSB galaxy distribution would allow to know if the predictions of biased galaxy formation models are correct.

IV. DWARF AND LSB GALAXY DISTRIBUTION

a) The redshift survey

The redshift survey was assembled from the Uppsala General Catalogue[5] of Galaxies which attempts to list all northern galaxies ($\delta \geq -2°30'$) visible on the Palomar Sky Survey with a blue diameter larger than 1 arcmin, with a Hubble type of Dwarf, Dwarf Irregular, Dwarf Sipiral, Ir, Sc-Irr or S-Irr; a de Vaucouleurs type of Sd-dm, Sdm, SDm or Im; or in the absence of these classification, a van den Bergh luminosity class of $IV - V$ or V. These selection criteria give a sample of 1845 galaxies. The HI detection rate is $\sim 84\%$, giving a total sample of 1557 galaxies with $21cm$ redshifts. A more detailed description of the observational sample and of the observing techniques can be found in Thuan[6] et al. (1990). For the statistical studies, we restrict the total sample to a galactic latitude $\|b\| \geq 40°$ to eliminate effects of galactic absorption and to $\delta \geq 0°$, because of the reduced sensitivity of the arecibo telescope at its extreme declination limit $-2° < \delta < 0°$. We also adopt a velocity cut-off of $10,000 km s^{-1}$ because of the large incompleteness for larger velocities. The statistical sample thus defined contains a total of 860 galaxies. In the same region of the sky ($\delta \geq 0°, \|b\| \geq 40°$) and with a velocity cut-off of $10,000 km s^{-1}$, there are 2154 CfA bright galaxies. It is first important to notice that, because the data under analysis has been obtained with a $21 - cm$ redshift survey, our results concern only HI-rich dwarf and LSB galaxies. Thus, we cannot put constaints on the spatial distribution of HI-deficient objects such as dwarf ellipticals.

Figures $1a - 1b$ shows the redshift distribution of the whole statistical sample for the dwarf-LSB galaxies and CfA bright galaxies. The shape of the two distributions

is very similar. We observe in the two distributions two peaks, one at 1000 km s^{-1} which corresponds to Virgo cluster, and one more pronounced for the bright galaxies at 5000 km s^{-1} which corresponds to Perseus Pisces supercluster. We present also (figures $2a - 2b$) two declination slices in which the dwarf-LSB galaxies are plotted as circles and the CfA bright galaxies as crosses. These figures seems already to indicate, dwarf-LSB galaxies follow very closely the structures delineated by the bright galaxies and do not fill in the voids seen in the bright galaxy distribution. But figures $3a - 3b$ show up a difficulty, our sample of dwarf-LSB galaxies consists in fact of at least two subsets: true low-luminosity dwarf galaxies and large, higher-luminosity LSB galaxies. We use the HI width Δv to distinguish between the two sub-populations. We define now the dwarf population as composed of galaxies with $\Delta v \leq 100$ km s^{-1} and the LSB population as composed of galaxies with $\Delta v > 100$ km s^{-1}. Figures $3a$ and $3b$ show respectively the redshift distribution of the dwarf and LSB populations. The dwarf velocity distribution is narrow and drops off sharply beyond ~ 2000 km s^{-1}, while the LSB velocity distribution is much broader, extending up to ~ 7000 km s^{-1}. In the following, we shall discuss, whenever possible, these two populations separately.
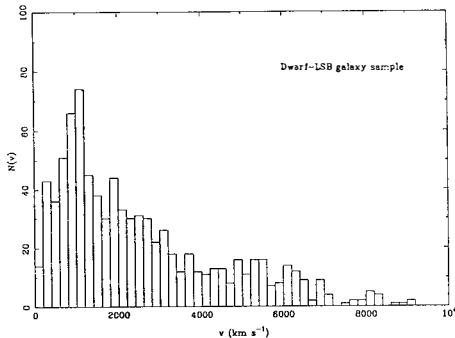

Figure 1a

Figure 1b

Figure 2a

$N_{CfA} = 226$
$N_{Dw} = 99$
$0 < \delta < 10$

Figure 2b

$N_{CfA} = 258$
$N_{Dw} = 82$
$10 < \delta < 20$

$\Delta v \leq 100$ km s^{-1}
$N_{Dw} = 253$

Figure 3a

$\Delta v > 100$ km s^{-1}
$N_{Dw} = 414$
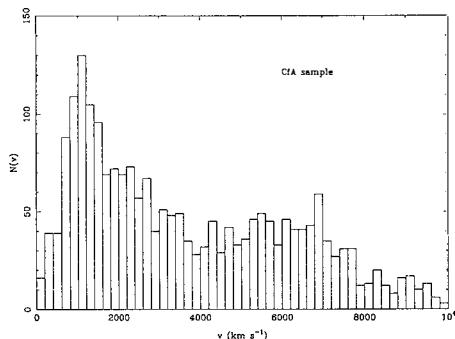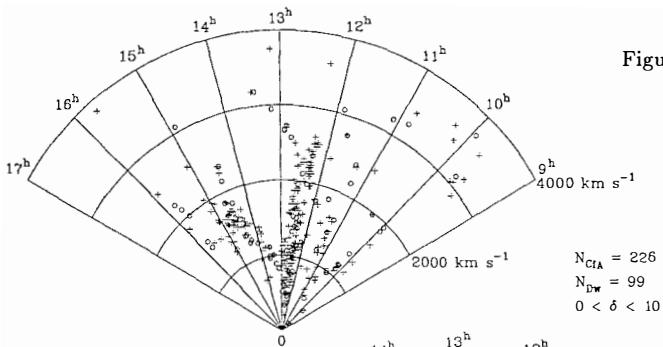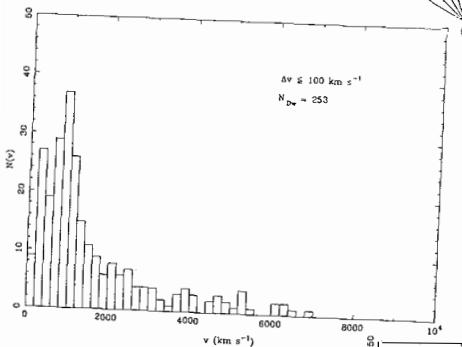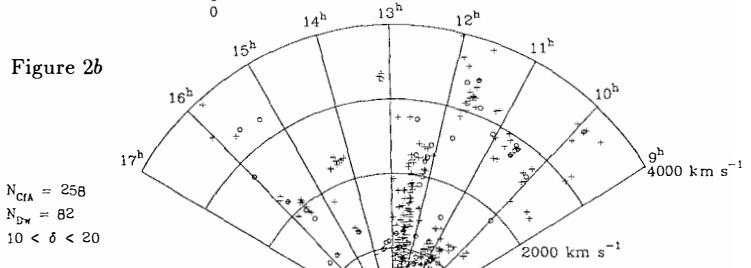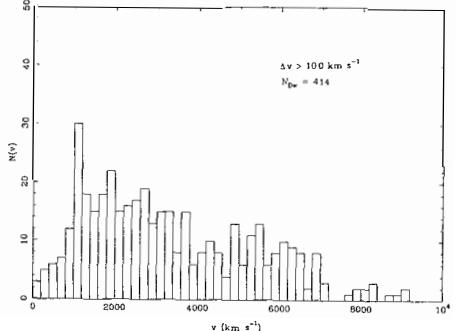
Figure 3b

b) The sub-samples

In this paper we only discuss the statistical analysis from completed volume limited sub-samples. We consider four sub-samples. We define two dwarfs samples, one (labeled as $D1$) which contains all dwarfs with a linear diameter d superior to 4 h^{-1} kpc, and the other one $(D2)$ with $d \geq 8$ h^{-1} kpc. With the constraint that the angular diameter be superior to 1 arcminute, this corresponds to a superior velocity limit of 1375 km s^{-1} for $D1$ and of 2750 km s^{-1} for $D2$. In the same manner, we define two volume-limited LSB samples with $d \geq 11.58$ h^{-1} kpc $(LSB1)$ and $d \geq 16$ h^{-1} kpc $(LSB2)$ corresponding respectively to the velocity limits of 3981 km s^{-1} 5500 km s^{-1}. Table 1 lists the observed number of galaxies and the expected number of galaxies in each sub-samples with and without Virgo Infall correction[34]. The expected number of galaxies is estimated from selection function[6], it is crucial in the estimation of the correlation function. Table 1 also presents bright CfA galaxies in the same four volumes.

TABLE 1: Completed volume-limited sub-samples for dwarf/LSB and Bright CfA galaxies.

			Dwarf/LSB				CfA			
			$V_i = 0$		$V_i = 220$		$V_i = 0$		$V_i = 220$	
Sample	d (h^{-1}kpc)	v (km s^{-1})	N_{obs}	N_{exp}	N_{obs}	N_{exp}	Nobs	Nexp	Nobs	Nexp
D1	$d \geq 4$ (M $\leq$ -16.2)	v $\leq$ 1375	179	99	129	89	340	169	236	147
D2	$d \geq 8$ (M $\leq$ -17.7)	v $\leq$ 2750	202	136	221	146	489	330	513	325
LSB1	$d \geq 11.58$ (M $\leq$ -18.5)	v $\leq$ 3981	157	130	207	157	421	407	516	430
LSB2	$d \geq 16$ (M $\leq$ -19.2)	v $\leq$ 5500	110	95	134	114	319	398	404	442

c) Results in redshift space

We now calculate the auto-correlation functions for dwarf and LSB galaxies in the complete sub-samples defined in Table 1 and that for bright CfA galaxies in the corresponding volumes. The calculation is done in redshift space, we assume that the cosmological redshift is proportional to distance and neglect peculiar motions. Table 2 gives the results of the least-square power-law fits of the form $\xi(s) = \left(\frac{s}{s_0}\right)^{-\gamma}$, to the auto-correlation functions, with the formal errors of the fits. In general, the data cannot be fitted by a single power law over the whole radius range, so the results presented in Table 2 refer to the $0.6 - 8.5$ h^{-1} Mpc range. The slopes of the auto-correlation functions for the Dwarf-LSB samples are similar to those for the bright galaxy samples. The correlation lenghts are systematically smaller for the Dwarf-LSB galaxies as compared to those for the bright galaxies.

TABLE 2: The two-point corelation function for dwarf/LSB and bright CfA galaxies in completed volume-limited samples in redshift space with Virgo infall correction (Vi = 220 km s-1)

	Dwarf/LSB		CfA	
Sample	γ	s_0(h^{-1} Mpc)	γ	s_0(h^{-1} Mpc)
D1	1.63 ± 0.09	4.29 ± 0.71	1.47 ± 0.05	6.55 ± 0.70
D2	1.35 ± 0.08	4.62 ± 0.76	1.32 ± 0.10	8.44 ± 1.95
LSB1	1.50 ± 0.17	5.25 ± 1.76	1.41 ± 0.08	7.50 ± 1.28
LSB2	1.65± 0.18	6.45 ± 2.16	1.66± 0.06	6.12 ± 0.74

TABLE 3: The cross-correlation function for dwarf/LSB and bright CfA galaxies in completed volume-limited samples in redshift space with Virgo infall correction (V_i = 220 km s^{-1}).

Sample	Dwarf/LSB		CfA	
	γ	s_0(h^{-1} Mpc)	γ	s_0(h^{-1} Mpc)
D1	1.56 ± 0.06	5.10 ± 0.48	1.64 ± 0.06	4.27 ± 0.41
D2	1..25 ± 0.10	6.24 ± 1.42	1.40 ± 0.07	5.44 ± 0.76
LSB1	1.46 ± 0.06	5.84 ± 0.83	1.24 ± 0.06	6.26 ± 0.85
LSB2	1.70 ± 0.12	4.51 ± 1.04	1.49 ± 0.07	5.83 ± 1.05

The cross-correlation function in redshift space between dwarf-LSB and bright galaxies is another way for comparing the spatial distribution of the two galaxy population. The results of the power-law fits to the cross-correlation functions are given in table 3. As a check, we have computed the cross-correlation functions in two ways, in one case counting the number of dwarfs around bright galaxies (ξ_{BD}) and in the other, counting the number of bright galaxies around dwarfs (ξ_{DB}). Within the errors, both methods give the same results, which shows that the normalization of the correlation function is done correctly. The correlation lenghts are intermediate between the dwarf-LSB and bright galaxy redshift auto-correlation lenghts. Note that we do not see, in any of the sub-samples analysed here, the effect discussed by Eder[32] et al (1989): a drop at small separations ($s \sim 0.75$ h^{-1} Mpc) in the amplitude of the dwarf-LSB galaxy auto-correlation function and of the dwarf/LSB-bright galaxy cross-correlation function, an effect which Eder[32] et al (1989) attribute to the tidal disruption of low-mass irregulars when they are too close to more massive bright galaxies. We think that the discrepancy is due to the fact that Eder[32] et al (1989) do not use the same estimator (see above §II) to measure the correlation function[14]. We show as examples, the correlation functions for Dwarf (open circles) and bright CfA galaxies (filled circles) in the $D1$ (figure 4).
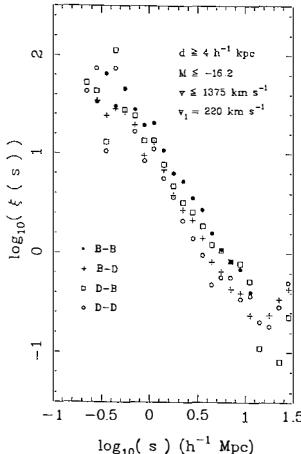

Figure 4

Comparison of Table 2 and 3 and examination of figure 4 show the existence of slight difference between the auto-correlation of Dwarf/LSB, those of bright galaxies and their cross-correlation functions. We would like to quantify these differences. For this we examine the ratio α_{DB} and its counterpart α_{BD} (see definition in section II). These ratios permit to examine possible difference between various galaxy populations, without being too sensitive to the uncertainties discussed in section II. α_{DB} and α_{BD} are both nearly constant with the distance s, independant of the Virgocentric correction and subsample. We obtain $\alpha_{D/LSB-B} \sim 0.7 \pm 0.1$ and $\alpha_{B-D/LSB} \sim 1.3 \pm 0.1$.

If we consider that the cross-correlation function ξ_{BD} and ξ_{DB} and the auto-correlation ξ_{BB} and ξ_{DD} can be written as, $\xi_{BD} = \xi_{DB} = b_B b_D \xi_{\rho\rho}$, $\xi_{BB} = b_B^2 \xi_{\rho\rho}$ and $\xi_{DD} = b_D^2 \xi_{\rho\rho}$. The quantity α_{DB} yields to the ratio of the bias parameter for dwarfs (b_D) to that for bright galaxies (b_B) as $b_D/b_B \sim 0.8$. If b_B is 2.5, the value needed in a cold dark matter scenario with $\Omega = 1$, then b_D is 2. This value is far enough from unity that it rules out certain classes of biased galaxy formation theories which predict an uniform spatial distribution of dwarf-LSB galaxies1).

d) Results in real space

Galaxies have in fact peculiar motions superposed on the Hubble expansion flow, and as a consequence, the redshift map is not a true space map: the galaxy positions are distorted along the line of sight. In order to assess the contribution of peculiar velocity to $\xi(s)$, we need to compute the two-dimensional correlation function $\xi(r_p, \pi)$, in distinguishing the separations perpendicular (r_p) and parallel (π) to the line of sight. Figure 5a show the function $\xi(r_p, \pi)$ for the complete sample $D2$. The contour $\xi(r_p, \pi) = 1, 2$ and 4 are shown. We show only the quadrant with positive π and r_p since the quadrant with negative π and positive r_p can be derived by mirror symmetry. The $\xi(r_p, \pi)$ contours are clearly elongated in the π direction, showing convincingly that dwarf-LSB galaxies possess significant peculiar velocities which distort the redshift maps as compared to the true spatial distribution maps. To obtain the space correlation function $\xi(r)$, we first compute the projected function $w(r_p)$ defined as $w(r_p) = \frac{1}{H_0} \int_0^{v_L} d\pi$ $\xi(r_p, \pi)$, where v_L is the upper limit for the redshift difference π. We have chosen v_L to be 1250 $km\ s-1$ for all samples. This is large enough to include nearly all correlated pairs and peculiar velocities and small enough so that $\xi(r_p, \pi)$ is not too noisy. Then we fit $w(r_p)$ to a power law model for $\xi(r)$. We show as example, the projected correlation function for Dwarf galaxies in the $D2$ (figure 5b). In the special case where $\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}$, $w(r_p) = A\ r_p^{1-\gamma}$ with $A = r_0^\gamma \Gamma(1/2) \Gamma[(\gamma - 1)/2]/\Gamma(\gamma/2)$. Thus a power law fit to $w(r_p)$ yields directly r_0 and γ. Table 5 lists the correlation lenght and slope of the spatial two-point correlation, function for Dwarf/LSB and bright CfA galaxies in completed volume-limited samples with Virgo Infall correction ($V_i = 220km\ s^{-1}$).

TABLE 4: The spatial two-point correlation function for dwarf/LSB and bright CfA galaxies in completed volume-limited samples with Virgo infall correction ($V_i = 220$ km s^{-1}).

Sample	Dwarf/LSB γ	s_0(h^{-1} Mpc)	CfA γ	s_0(h^{-1} Mpc)
D1	1.40 ± 0.12	4.26 ± 0.31	1.67 ± 0.09	6.27 ± 0.29
D2	1.62 ± 0.07	3.95 ± 0.16	1.66 ± 0.06	6.68 ± 0.20
LSB1	1.72 ± 0.11	3.65 ± 0.21	1.61 ± 0.03	5.90 ± 0.09
LSB2	1.60 ± 0.30	4.51 ± 0.75	1.94 ± 0.04	5.16 ± 0.08
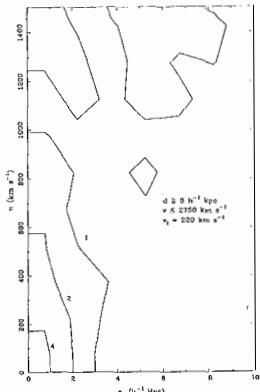
Figure 5a
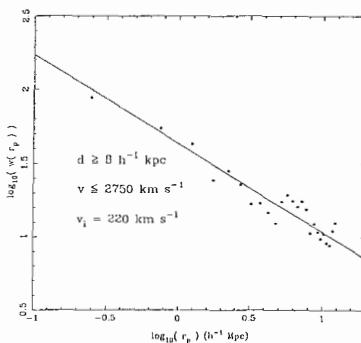
Figure 5b

References

1. Dekel, A., Silk, J. 1986, *Ap. J.* **303**, 39.
2. Davis, M., Djorgovski, S. 1985, *Ap. J.* **299**, 15.
3. Bothun, G.D., Beers, T.C., Mould, J.R., and Huchra, J.P. 1986, *Ap. J.* **308**, 510.
4. Thuan, T.X., Gott, J.R., and Schneider, S.E. 1986, *Ap. J. (Letters)* **315**, L93.
5. Nilson, P. 1973, Uppsala Astr. Obs. Ann., Vol 6 (Uppsala General Catalog of Galaxies).
6. Thuan, T.X., Alimi, J.-M., Gott, J.R., and Schneider, S.E. 1990, submitted to *Ap J.*
7. Shandarin, S.F., Zeldovich, Ya.B. 1986, 2nd *ESO-CERN Symposium.*
8. White, S.D.M. 1979, *M. N. R. A. S.*, **186**, 145.
9. Balian, R., Schaeffer, R. 1989 *Astr. Ap.*, **XXX**, XXX.
10. Alimi, J.-M., Blanchard, A., and Schaeffer, R. 1990 *Ap. J. (Letters)* **349**, L5.
11. Peebles, P.J.E. 1980 *The Large Scale Structure in the Universe*, Princeton University Press.
12. Davis, M., Peebles, P.J.E. 1983 *Ap. J.* **267**, 465.
13. Rivolo, A.R. 1986 *Ap. J.* **301**, 70.
14. Blanchard, A., Alimi, J.-M. 1988 *Astr. Ap.*, **206**, L11.
15. Einasto, J., Klypin, A.A., and Saar, E. 1986 *M. N. R. A. S.*, **219**, 457.

16. de Lapparent, V., Geller, M.J., and Huchra, J.P. 1988 *Ap. J.* **332**, 44.

17. Efstathiou, G., Ellis, R.S., and Peterson, B.A. 1988 *M. N. R. A. S.,* **232**, 431.

18. Kaiser 1987 *M. N. R. A. S.,* **227**, 1.

19. Groth, E., Peebles, P.J.E. 1977 *Ap. J.* **217**, 385.

20. Alimi, J.-M., Valls-Gabaud, D., and Blanchard, A. 1988 *Astr. Ap.,* **206**, L11.

21. Blumenthal,G.R., Faber, S.M., Primack, J.R., and Rees, M.J. 1984 *Nature* **311**, 517.

22. Rees, M.J. 1985 *M. N. R. A. S.,* **213**, 75p.

23. Dekel, A., Rees, M.J. 1987 *Nature* **326**, 455.

24. White, S.D.M., Davis, M., Efstathiou, G., and Frenk C.S. 1987 *Nature* **330**, 451.

25. Schaeffer, R. 1987 *Astr. Ap.,* **180**, L5.

26. West, M.J., Richstone, D.O. 1988 *Ap. J.* **335**, 532.

27. Davis, M., Efstathiou, G., Frenk C.S., and White, S.D.M. 1985 *Ap. J.* **292**, 371.

28. Bardeen, J.M., Bond, J.R., Kaiser, N., and Szalay, A.S. 1986 *Ap. J.* **304**, 15.

29. Davis, M., Djorgovski, S. 1985 *Ap. J.* **299**, 15.

30. Bothun, G.D., Beers, T.G., Mould, J.R., and Huchra, J.P. 1986 *Ap. J.* **308**, 510.

31. Giovanelli, R., Haynes, M.P., and Chincarini, G.L. (1986) *Ap. J.* **300**, 77.

32. Eder, J.A., Schombert, J.M., Dekel, A., and Oemler, A. 1989 *Ap. J.* **340**, 29.

33. Bingelli, B., Tarenghi, M. and Sandage, A. 1989 *Astr. Ap.,* **228**, 42.

34. Kraan-Korteweg, R.C., 1986 *Astr. Ap. Suppl. Ser.,* **66**, 255.

DARK MATTER SEARCH: THE SACLAY PROGRAM

O.Besida, G.Chardin, G.Gerbier, E.Lesquoy,
J.Rich, M.Spiro, C.Tao*,D.Yvon, S.Zylberajch,
presented by C.Tao
DPhPE,CEN Saclay, France

Abstract

This report summarizes the current efforts by the Saclay group concerning direct Dark Matter search. The emphasis in the last two years has been put on the relatively "easier" cosmion detection (picobarn or larger cross-sections). The silicon experiment we (LBL /Berkeley /UCSB /Saclay Collaboration) performed rules out large regions in mass and cross-section for would- be cosmions with coherent vector couplings. Developments on some new detectors which could be used in the search for cosmions with axial couplings are sketched. Possible experiments searching for other dark matter candidates are briefly described to complete the picture of the variety of interests and efforts carried within our group.

---

1 Introduction

Cosmions have been invented to solve simultaneously the problem of Dark Matter and the Solar Neutrino Puzzle. There is an ongoing debate within the astrophysical community on the compatibility of the Cosmion hypothesis with observations. I refer the reader to the talk of F.Martin de Volnay in these proceedings [1](and references therein) for more details. Cosmions are WIMPs (Weakly Interacting Massive Particles) which would form the Dark Matter Halo of our galaxy. They would be gravitationnally trapped in the core of the Sun, transport energy from the inner to the outer regions, cool down the center, thus lowering the Beryllium and Boron neutrino production, and barely affect the important pp neutrino rate. The SAGE and GALLEX experiments will provide in the next years critical and fundamental results about the solar neutrino flux. They are however insensitive to the Cosmion mechanisms which affect the more central regions of the Sun.

Direct detection of Cosmions [2] would be a major contribution to this debate. As the earth moves through the galactic sea of dark matter, the particles, if they exist, scatter from nuclei with a recoil energy T which is given by

$$T = Mm^2v^2(1 - cos\theta)/(m + M)^2$$

for nuclear mass M, cosmion mass m and c.m. scattering angle θ. For GeV cosmions, T is of the order of one keV.

Recent work has been performed to set more precise constraints on Cosmions using a solar evolution code [1]. With the picobarn cross-sections, and masses of a few GeV, which they must have to affect the solar neutrino flux and respect the existing known parameters of the Sun, Cosmions do not belong to the zoo of known or commonly accepted particles: the LEP results have ruled out the existence of a fourth family of neutral particles coupled to the Z_0, thus excluding many of the ad-hoc Cosmion models constructed by theorists in the last years. As for neutralinos, in the minimal supersymmetric schemes, they are now expected to be of masses above 30 GeV, much too heavy to solve the solar neutrino puzzle. It is now commonly accepted that a positive detection of Cosmions would have dramatic consequences on particle physics.

2 Vector Coupling Cosmions and the Silicon experiment

In these models, not only are Cosmion cross-sections about 100 times the weak cross-sections, but cross-sections on nuclei increase roughly by the square of the number of nucleons times the reduced mass factor. Such particles more massive than 9 GeV/c^2 have already been excluded by Germanium experiments, and the interest of replacing Germanium by a Silicon array is to lower the mass limits (down to 3 GeV/c^2).

2.1 The Silicon set up at the Oroville Dam

The LBL/ UCSB/ Berkeley/ Saclay collaboration initiated two years ago based on idea described in references [3]; we use the double beta Germanium apparatus placed under 600 m.w.e. overburden in the powerhouse of the Oroville dam in California, replacing the Germanium detector by an array of pure Silicon detectors. Two of the original 8 Germanium detectors of 0.9 kg each remain with the Si array inside a cavity formed by ten blocks of NaI of 15 cm thickness. This anticoincidence shield with a threshold of 30 keV is in turn inside a very pure lead shield of 20 cm thickness.

As it proved difficult to find very pure Silicon detectors mined underground and protected throughout fabrication and transport from cosmic ray interactions, it was decided to start a test run with unselected Si with relatively large backgrounds. The digitised pulse heights and arrival times of the signals from each of the four Si detectors, the remaining Germanium detectors, and each NaI scintillator are recorded for off-line analysis. One Si detector has both a lower threshold and less background than the others, and the data from that unique detector is reported here and used to set limits. Its energy spectrum, which has been normalised in energy by pulser information checked against known photopeaks is shown in figure 1. The gaussian electronic noise ($\sigma = 0.22$ keV) gives an effective threshold of about 1.1 keV, which is about a factor 3 lower than the threshold obtained with Ge. The background, however, with this unselected Si material, is two orders of magnitude higher than what one measures with the best Ge detectors. The background includes β decay spectra from Tritium (18.6 keV, 12.3 y of halflife), ^{32}Si (225 keV, 104 y halflife), and ^{210}Pb (63.1 keV, 22.3 y halflife). Tritium is a spallation product produced in the Si by cosmic rays. The presence of ^{32}Si at a rate of about 300 counts/kg/day indicates that the the detector was obtained from surface sand as a starting material, since ^{32}Si is produced in the atmosphere by cosmic rays interacting with Argon, and falls to the surface of the earth. Silicon obtained from a deep mine would not contain ^{32}Si.

2.2 Calibration of Si nuclear recoil

Slow recoiling nuclei are expected to produce less ionisation than a Compton or photoelectron of the same energy. In the sixties, neutron nuclei elastic scattering was used to calibrate the response of Germanium (C.Chasman[5]) and Silicon (A.R. Sattler[6]) detectors. As the measurements did not extend down to the keV region interesting for Cosmion searches, we have performed an experiment where we have observed Silicon recoils between 3.2 and 21 keV. The 4 MeV Van de Graaf of the Centre d' Etudes Nucléaires de Bruyères -le Châtel produces proton pulses of duration 2 ns every 1 μs. The time averaged current is typically 2 μA. Neutrons are produced via the reaction ^7Li (p,n) ^7Be (threshold of 1.880 MeV). For a given angle between the neutron and proton directions, the kinematics of the (p,n) reaction fix the neutron energy. Most of our measurements were made with

a proton beam of energy 1980 keV and a neutron target (the Si detector) placed at 30 degrees with respect to the proton beam, yielding a mean neutron energy of 200 keV.

The known energy of ^{55}Fe X-rays (5.9 keV) and ^{241}Am (13.9, 17.6, 59.5 keV) are used to establish the energy scale. Since nuclear recoils do not produce the same ionisation as an X-ray of the same energy, we refer to the measured energy on the scale established by X-rays as the "equivalent electron energy"(e.e.e.)

The energy resolution, as measured with X-rays from ^{55} and ^{241}Am sources is 410 eV (e.e.e.) (FWHM). The linearity of the system and trigger efficiency are checked with a pulser system.

The scattering angle of neutrons is varied to select a silicon recoil energy as determined by the (p,n) and elastic scattering kinematics. The trigger is a coincidence between the Si(Li) detector and the scintillation counter. For each event, we record the Si(Li) pulse height, the time interval between the scintillator and Si(Li) pulses, and the time interval between the scintillator and the proton beam pulse as determined by an inductive pickup coil.

The signal due to neutrons scattered off Silicon nuclei and detected in the scintillation counters appears clearly above the background, at a time which does correspond to the expected time of flight. After background substraction, the signal peaks are fitted with a Gaussian to determine the mean and width of the observed energy distribution.

Figure 2 shows the ratio between the observed energy and the calculated recoil energy as a function of the recoil energy for the 8 energies we measured. Also shown are the 2 lowest points obtained by Sattler. The superimposed curve is taken from Lindhard et al.[7]. There is reasonable agreement between the theoretical extrapolations and the observed values (Lindhard et al. state that the assumptions made in their model may lead to errors of order 10%). The ratio between the ionisation produced by a recoil Silicon nucleus and an electron ranges from 0.25 at 3.4 keV up to 0.4 at 22 keV. Details on the experiment can be found in [4].

2.3 Limits on vector coupling Cosmions

To place limits on the Cosmion mass and cross-section on silicon, the measured energy spectrum is compared with that expected from Cosmions. We take a halo density of 0.3 GeV/c^2/cm^3 or 5 x 10^{-25}g/cm^3 and a maxwellian cosmion velocity distribution with a root mean square velocity between 260 and 300 km/s and a velocity of the earth of 230 km/s. Those parameters reflect current estimates and the effect of their uncertainties on the results will be discussed below.

In figure 3, we plot the exclusion region for coherent vector coupling Cosmions. Cosmions masses and cross-sections on Silicon above and to the right of the straight (260 km/s) or dashed-dotted (300 km/s) lines are excluded. The exclusion contour is obtained by two different procedures using maximum likelihood methods. One involves fits the region -1.1 to 1.5 keV- most sensitive to light Cosmions with a straight line, then adds a dark matter signal to the fit (allowing the parameters

of the "background" line to change) until the fit can be rejected at a confidence level of 95%. The other procedure uses a large region (up to 225 keV) of the data, includes shapes of known backgrounds (such as the electronic noise and radioactivities described above) plus a quadratic polynomial and an exponential. The background model is required to give a χ^2 compatible with the number of degrees of freedom. Again, for each assumed Cosmion mass, a signal is added until the fit can be rejected at the 95% confidence level. The final conclusions of the two analyses are in close agreement.

For the assumptions given above, we plot the 2 SNU and 2.5 SNU curves obtained with the Saclay standard solar evolution code described by F.Martin de Volnay in these proceedings[1]. Cosmions with Coherent vector coupling become very unlikely, but are still not completely ruled out, especially due to the large uncertainties in the dark matter halo parameters.

3 Axial Coupling Cosmions and the status of a Hydrogen TPC

For particles with spin-dependent couplings, suitable detectors do not exist yet. We have proposed[9] and are currently studying the use of a low pressure hydrogen Time Projection Chamber (TPC) in a magnetic field (cf figure 4).

A low pressure TPC provides the imaging capacity for keV tracks. The magnetic field can be adjusted so that keV electrons spiral while keV protons leave a rather straight track.

A TPC solution would allow the measurement of the direction of recoil protons. An observed correlation with the direction opposite to the Earth's movement relative to the center of the galaxy would be a signature for Dark Matter which is an asset that no other foreseen detector can present.

A TPC as small as 1 m^3 could be sensitive to cosmions with an interaction cross-section of 4 pb or more at the level of 1 event/day, if the background can be kept low enough.

3.1 Technical feasibility of a hydrogen low pressure TPC

The average kinetic energy of protons recoiling from collisions with halo particles is around 2 keV. Proportional counters filled with hydrogen and methane are commonly used to measure neutron fluxes of energy between 1 keV and 1 MeV in experiments of material testing for breeder reactors. Protons down to 2 keV do produce electron-ion pairs in hydrogen with the same efficiency as electrons and higher energy protons (about 30 eV /ion-electron pair).

The total range of a 2 keV proton in hydrogen is less than 1 mm at normal pressure. In order to obtain a track of a reasonable length (few cm) a low pressure chamber is necessary. Typically at a pressure of 20 torr, 1 keV protons have a track length of 3 cm in hydrogen and produce 15 to 20 electrons.

3.2 Tests with a prototype

A.Breskin, "the world expert" in low pressure chambers, convinced us that it is possible to detect a single electron with a low pressure chamber. He even offered one prototype which was no longer of use to his group. We first checked that we could detect one single photoelectron produced on photocathodes by a pulsed UV lamp.

A series of measurements convinced us that we need a second stage of amplification in order to reach a reasonable gain and stable conditions of work. This was implemented at the end of 1989. With this set up, we measured longitudinal and transverse diffusions for different conditions in ethane and methane, and verified the results by Breskin et al, obtained previously. We are now proceeding to measurements with hydrogen and hydrogen mixtures and results should appear soon. We built in parallel a small magnet (1kgauss) in which we installed the chamber in order to check the spiralisation of electrons in a magnetic field. Indeed the measured transverse diffusions after 10 cm of drift are reduced according to expected factors.

We have equipped the chamber with existing electronics provided by Breskin and installed it with the magnet in the neutron beam of Bruyères-le Châtel for a one-week run in April. The multistep TPC is a very sensitive device and has difficulty operating in a large background environment (every single electron is multiplied by 10^6 or more). We have managed to shield the detector well enough that it could run reliably with standard beam luminosities.

Data is being analysed and we are hoping that the foreseen improvements (with FADC, better shielding and better alignment) will provide us in the next months the answers necessary for deciding if a 1 m^3 detector for cosmion search is feasible, ie, whether a proton of 1 or 2 keV gives a recognizable track, or not.

3.3 Background

An important effort on the choice of low radioactive material has been carried out by the double β experiments. Many groups have built low radioactive drift chambers. Their experience is important for the final design of a Dark Matter TPC. The Orsay LAL group is helping us designing and building a series of Geiger tubes with which we plan to measure the rates of background and if possible identify their different origins. The test is planned for the fall 1990 in the Fréjus Tunnel at the Modane Underground Laboratory (LSM).

4 Other Saclay Projects

4.1 NaI

NaI is cheap, exists already as veto in many underground experiments in large quantities (above 30 kg); last but not least, the Na and I nuclei have non integer

spins. We have investigated the potential of NaI as a Dark Matter detector[1].
The major problem is that the calculations for the rates of axial couplings are not
trivial. The spin of nuclei (and not the sum of the orbital and spin components) is
the value of interest to us but its value depends on the quark content of the nuclei.
A few predictions exist but there is an ongoing controversy[11]. To compare one
nucleus to another in the choice of detectors depends on the models.

On the experimental side, we realised many interesting points: First, the single
photo electron threshold can be as low as 300 eV; the rates of the NaI veto in the
GALLEX experiment are low and the light response of proton recoils had been
measured down to 100 keV. We have performed a calibration with the neutron
beam and the preliminary results are very positive. Not only do Na recoil nu-
clei yield light down to energies of a few keV, but about one third of the recoil
energy is turned into scintillation. Furthermore we have observed different pulse
shapes between 14 keV γ rays and 14 keV Na recoils, suggesting the possibility
for background rejections in these low energy regions. More details are provided
in reference[8].

4.2   Bolometers

Bolometers up to 280 g have been built with energy resolutions which approach
what is needed for Dark Matter searches (60 keV). It can be hoped that within a
few years, bolometers become good Dark Matter detectors. Since two of the fol-
lowing papers of this workshop discuss recent bolometer developments at length,
I only want to mention here that Saclay has started work in this field about 18
months ago and the group (Chardin and Chapellier essentially) has successfully
observed 1% resolution on α from an ^{241}Am source and γ above a threshold of 10
keV[10].

We are planning a calibration run with a cryostat and a saphire or silicon bolome-
ter in a neutron beam for the end of June as well as the construction of a low
radioactive cryostat for measurements in the Modane underground Laboratory.

4.3   CHAMPS

CHAMPS (CHArged Massive Particles) have been examined by DeRujula,
Glashow and Sarid[12] as eventual Dark Matter candidates. Their conclusion
is that a window of charged particles is left open in the region 10 to 1000 TeV at
the level of 10^{-11} [10 TeV/M].

Previously Fayet had suggested that supersymmetric particles would show up as
heavy isotopes of hydrogen and this had triggered a ENS /Institut Pasteur /Saclay
collaboration[13] looking for heavy isotopes of hydrogen in deep sea water. Sam-
ples of sea water from all world oceans have been extracted and centrifugated.
Water is passed through an Uranium sieve to reduce it into hydrogen and a laser

---

[1]R. Plaga first suggested this idea during a GALLEX collaboration meeting

478

separates hydrogen from its heavy isotopes. The absence of signal at twice the isotopic shift of Deuterium indicates the absence of a very heavy isotope.

The experiment started in 1987 and yields results better than 10^{-15} [10 TeV/M], which already excludes the existence of CHAMPS as Dark Matter candidates.

4.4   Brown Dwarves

Macroscopic Compact Halo Objects (MACHOS a term coined recently by Kim Griest) are a favourite among some astronomers as Dark Matter candidate. B. Paczynski[14] suggested in 1986 that monitoring of the light intensity of one million stars in the Large Magellanic Clouds would allow the detection of a Brown Dwarf (or any MACHO of the right mass) if they make up the Dark Matter of our galaxy. The MACHO would act as a gravitational lens and the resulting observed light from a distant star would be the sum of the deflected intensities. Saclay has initiated a collaboration between astrophysicists and particle physicists for such a search. The reader should refer to the paper of M. Moniez in these proceedings for details on the proposal.

5   Acknowledgements

References

[1] " WIMPs and the Saclay Solar evolution code", F. Martin de Volnay (et al.), communication in these proceedings.
also, Y.Giraud-Héraud, J.Kaplan, F.Martin de Volnay, C.Tao, S.Turck-Chièze, in Proc. of the Int.Conf. "Inside the Sun", ed. G.Berthomieu, M.Cribier (Versailles, France, 1989)

[2] M.W. Goodman and E. Witten, Phys. Rev. D31 (1985) 3059.

[3] B.Sadoulet, J.Rich, M.Spiro, and D.O.Caldwell, Ap.J.324, L75 (1988). The paper with results presented here is in the final stage of discussion and will be submitted to Phys.Rev.Lett.

[4] G. Gerbier et al.,CEA Saclay preprint, DPhPE 90/02, to be submitted to Phys.Rev.Lett. together with the Silicon results above.

[5] C.Chasman, K.W.Jones, R.A.Ristinen, P.R.L. 15 (1965) 245.

[6] A.R.Sattler, Phys.Rev. 138 6A (1965) 1815.

[7] J.Lindhard, V.Nielsen, M.Scharff, P.V.Thomsen, Mat.Fys. Medd. Dan. Vid. Selsk. 33, no.10 (1963).

[8] G.Gerbier, in Proceedings of Moriond Workshop, January 1990.

[9] G.Gerbier,J.Rich,M.Spiro and C.Tao, Nucl.Ph.B 13 (1990) 207.

[10] M.Chapellier et al., to appear in proceedings of the third Workshop on low temperature detectors, Gran Sasso 1989.

[11] L.Krauss, preprint in preparation.

[12] A.De Rujula, S.Glashow, U.Sarid, preprint HUTP-89/A001 and CERN TH-5214/89 to appear in Nucl.Physics.

[13] B.Pichard et al, Phys. Lett.B 193 (1987) 383.

[14] B.Paczynski, Ap.J. 304 (1986)1.
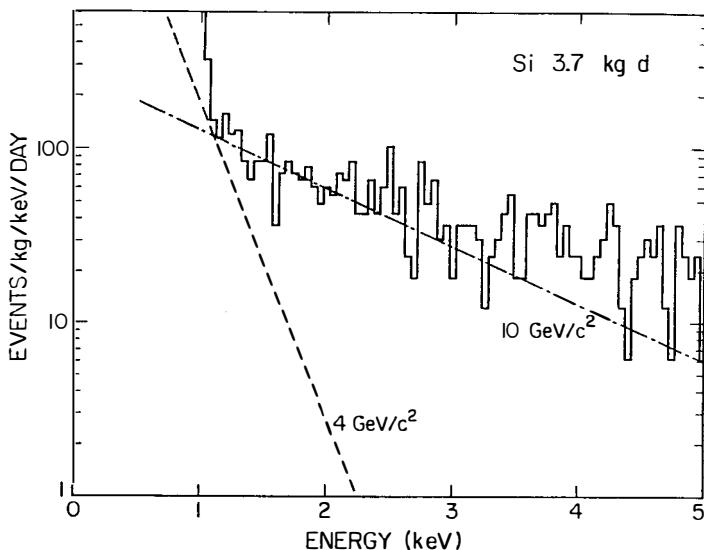
Figure 1 : Low energy region of the ionization detected in the 17 g Si(Li) detector for 3.7 kg.days of data. Also shown are the signals expected from 4 and 10 GeV/c² Cosmions with corresponding cross sections on the exclusion contour (v_{rms} = 300 km/s) of Figure 3.
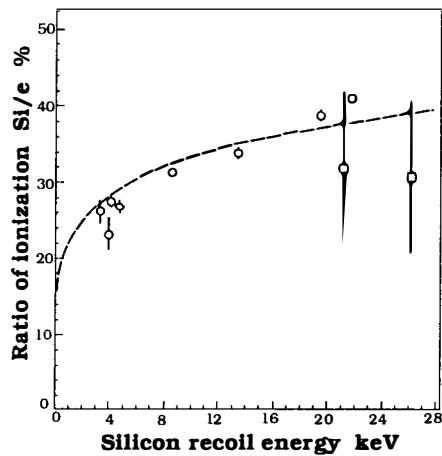

Figure 2: Ratio between the ionisation produced by a Si recoil and the ionisation produced by an electron of the same kinetic energy as a function of the kinetic energy. Circles are data points from the present experiment. Squares are from Sattler [9], the curve represent the prediction by Lindhard et al. [11].
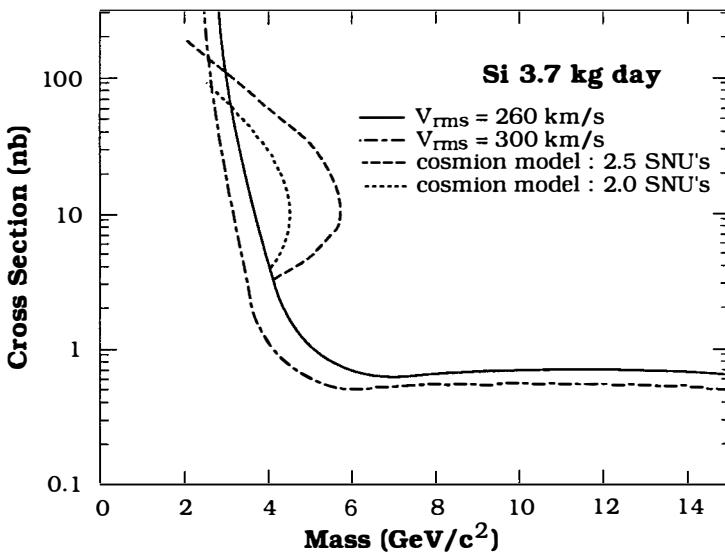
Figure 3 : Exclusion plots for two values of the velocity of Dark Matter particles as functions of their mass and elastic cross section on Silicon. Also shown are the expected curves where should lie the Cosmions for a resulting neutrino flux of 2 and 2.5 SNU's [2].
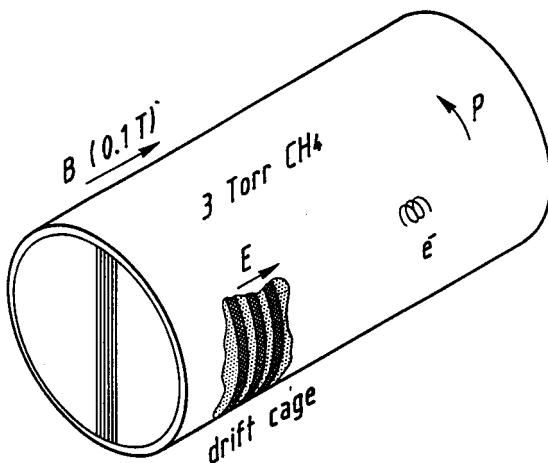

Figure 4: A sketch of a possible low pressure hydrogenous Time Projection Chamber.

THE QUEST FOR Ω: COMPARISON OF
DENSITY AND PECULIAR VELOCITY FIELDS

Amos Yahil

State University of New York at Stony Brook

ABSTRACT

The density field traced by the *IRAS* galaxies is compared with the smoothed peculiar velocity map generated by the *POTENT* scheme of Bertschinger and Dekel. In regions in which both datasets are reliable, the fields are found to be proportional to each other, as expected in gravitational instability theory, with the constant of proportionality in the range $\lambda \equiv f(\Omega)/b = 0.75 - 1.15$. If the *IRAS* galaxies are unbiased tracers of the mass, $b = 1$, then $\Omega = 0.6 - 1.25$, but the result can also be be interpreted as a measurement of b for a given value of Ω.

Previous, lower, estimates of Ω are re-examined to determine the severity of the contradiction. The Virgocentric infall model is rejected, primarily because of its failure to take into account the local void in the North Supergalactic Hemisphere. None of the other differences seem irreconcilable. Further work is outlined on clusters of galaxies, the "Cosmic Virial Theorem", N-body simulations, morphological segregation, and the universality of the relation between density and velocity.

1. INTRODUCTION

A basic paradigm in current cosmological research is the assumption that the present structure and motions in the universe are due to the growth of gravitational instabilities from an initially highly homogeneous background. In particular, initial peculiar velocities (deviations from the smooth Hubble flow) are adiabatically damped by the expansion of the universe, and the peculiar velocity of any galaxy today is due to the gravitational pull of the matter around it, and should thus be mirrored in the distribution of this matter.

In linear perturbation theory, the peculiar velocity of any galaxy in space is directly proportional to the *current* dipole moment of the matter distribution around it. Specifically, the peculiar velocity is given by (Peebles 1976a):

$$\mathbf{V}(\mathbf{r}) = \frac{H_0 f(\Omega)}{4\pi} \int \delta(\mathbf{r}') \frac{(\mathbf{r}' - \mathbf{r})}{|\mathbf{r}' - \mathbf{r}|^3} d^3\mathbf{r}' \quad , \tag{1}$$

where H_0 is the Hubble Constant, $\delta(\mathbf{r}') = (\rho(\mathbf{r}') - \rho_0)/\rho_0$ is the *mass* density expressed in units of the average density ρ_0, and

$$f(\Omega) = d \ln\delta / d \ln a \approx \Omega^{0.6} \quad , \tag{2}$$

is the logarithmic derivative of the amplitude of the growing mode of density perturbations, with respect to the scale factor a. Curvature effects are neglected; it is assumed that the dominant acceleration in Eq. (1) is due to material on scales much smaller than the horizon.

Peculiar velocities have been measured in two ways. The observed dipole anisotropy of the cosmic microwave background (CMB), when corrected to the barycenter of the Local Group (LG), is interpreted as due to the latter's motion (e.g., Lubin and Villela 1986). Additionally, many galaxies have had their distances estimated independently of their redshifts, using the infrared Tully-Fisher relation for spiral galaxies (Aaronson *et al.* 1989; and references therein) and the $D_n - \sigma$ relation for ellipticals (Lynden-Bell *et al.* 1988; and references therein); these provide estimates of their peculiar velocities.

If a set of galaxies trace the mass, at least when averaged on the scales over which density is to be resolved, then the mass density and the galaxian number (or luminosity) density are proportional. Note that the constant of proportionality is not an issue; as seen in Eq. (1) only *relative* densities are needed, and $\delta(\mathbf{r}')$ can be taken to refer to the galaxian density. The combined measurement of peculiar velocity and galaxian density fields thus opens up the exciting possibility of measuring a dynamic value of Ω.

If the galaxies used do not trace the mass, i.e., if galaxy formation is biased, then a prescription is needed for converting mass density into number density. (See Dekel and Rees 1987 for a review of biasing.) In the linear biasing model, the perturbation in number density is a constant, b, times the mass perturbation. Eq. (1) therefore continues to hold, provided that $f(\Omega)$ is replaced by $\lambda \equiv f(\Omega)/b$. In this approximation, therefore, it is not possible to break the parametric degeneracy between Ω and b, and the comparison of the density and velocity fields only yields the combination λ; nonlinear effects are needed to break the

degeneracy. In this paper I concentrate on the linear approximation, and determine λ. I usually interpret this as a measurement of Ω, assuming that $b = 1$, but an interpretation in terms of b, assuming, for example, that $\Omega = 1$, is equally valid.

The number of galaxies involved in the definitions of the velocity and density fields is in each case already in excess of a thousand, so the comparison between the two fields is an overconstrained problem. It can therefore also serve to test the basic paradigm that galaxies trace the mass that has given rise to their peculiar velocities. Gravitational instability theory predicts a *universal* relation between mass density and velocity, which is only a function of the underlying cosmology. As long as galaxy formation, even if biased, is a universal process, there is a one-to-one correspondence between mass density and galaxian density, and hence between the observed velocity and galaxian density fields. In competing theories, on the other hand, there are usually other local factors which affect the relation between the two, destroying the universality of the relation between them.

Previous mappings of the local density structure, based on optically selected galaxies (Yahil, Sandage, and Tammann 1980; Davis and Huchra 1982; da Costa *et al.* 1988; Tully and Fisher 1987), have been constrained in sky coverage and depth. This paper presents the results of a complete redshift survey of galaxies detected by the *Infrared Astronomical Satellite* (*IRAS*),[1] undertaken by M. Strauss, M. Davis, J. Huchra, J. Tonry and A. Yahil. We have obtained redshifts for a complete sample of 2649 galaxies, flux-limited to 1.936 Jy at $60\mu m$, which is selected over 11.06 steradians of the sky. The sample selection and biases inherent in the *IRAS* database are discussed in Strauss *et al.* (1990).

The structure of the paper is as follows. The methodology involved in converting a redshift survey into a density map is briefly reviewed in § 2, and the comparison with velocity fields is discussed in § 3. The essence of the paper is the comparison, § 4, of the *IRAS* density field with the observed velocity field, smoothed by the *POTENT* procedure, reviewed in this volume by Bertschinger (1990). The resulting estimate of Ω is compared with previous, lower, values in § 5.

2. RECIPE FOR 3–D DENSITY MAPS

The transformation of a redshift survey to a 3–d density map involves a number of crucial ingredients. Full details are given by Yahil, Strauss, and Davis (1990). The main points are:

1. *Selection.* The selection algorithm of the galaxies must be precisely known, and the sample must be complete according to this algorithm. Typically one has a sample limited by flux, angular diameter, or both. For the sake of concreteness I refer to the *IRAS* survey, which is flux limited.

---

[1] The Infrared Astronomical Satellite was developed and operated by the U.S. National Aeronautics and Space Administration (NASA), the Netherlands Agency for Aerospace Programs (NIVR), and the U.K. Science and Engineering Research Council (SERC).

2. *Extinction.* Galactic extinction should be minimal, so the flux limit can be homogeneously applied.

3. *Selection Function.* The fraction of galaxies observed is a decreasing function of distance, because a smaller fraction of the luminosity function falls above the flux limit at greater distances. In order to convert the raw number counts into true densities, it is therefore necessary to weight each observed galaxy in inverse proportion to the selection function, the fraction of the luminosity function included in the sample at each distance. This procedure is only valid if the luminosity function is *universal* (Yahil *et al.* 1980), i.e., if the relative densities of galaxies of different luminosities are everywhere the same, and the normalization of the luminosity function is a local measure of density. With this assumption, the selection function depends only on distance from us, and the weighting scheme can be employed until statistical noise becomes the limiting factor at large distances. In this paper I use this procedure to a distance of 8000 km s^{-1}.[2]

4. *Transformation from Redshift to Configuration Space.* We seek to know the density structure in real (configuration) space. But the redshift of each galaxy (in the LG reference frame) is a sum of its Hubble recession velocity and its peculiar velocity relative to the Local Group, projected along the line of sight.

$$cz = H_0 r + \hat{\mathbf{r}} \cdot (\mathbf{V} - \mathbf{V}_{LG}) \quad . \tag{3}$$

Here $\hat{\mathbf{r}}$ is the unit vector in the direction of the galaxy, and V and $\mathbf{V}_{LG}$ are the peculiar velocity vectors of the galaxy and the LG, respectively. Gravitational instability theory, in either its linear form, or in a nonlinear variant, is used to transform from redshift to configuration space. For example, in the linear case self-consistent solutions are sought to Eqs. (1) and (3) for all the galaxies, given their positions and redshifts.

5. *Adequate Sky Coverage.* Since gravity is nonlocal, such self-consistent solutions can only be found for samples of galaxies that adequately survey the density structures giving rise to the gravitational field, both at the positions of the galaxies and at the Local Group. This requires adequate sky coverage; surveys which do not satisfy this criterion, such as optical surveys which exclude a significant zone of avoidance around the Galactic plane, need to make use of the peculiar velocity field determined from an all-sky survey.

6. *Excluded Zones.* Even in surveys which cover essentially the entire sky, some part of the sky is not surveyed, e.g., the *IRAS* survey covers 88% of the sky. It is necessary to account for the gravity due to the excluded areas; in practice artificial "galaxies" are introduced into these regions. The simplest law, according to which the artificial galaxies are distributed,

---

[2] The relative density field, δ, and the velocity field $\mathbf{V}$ are both independent of the extragalactic distance scale. To avoid confusion, I therefore always quote distances in km s^{-1}. The Hubble constant is introduced for the sake of concreteness when necessary, but it cancels out in the final results for relative density or velocity.

assumes that the density there shows no structure. If the unsurveyed region is narrow enough, though, it is also possible to interpolate the density from either side.

7. *Verification.* The above steps result in a computed density field. It remains to verify *a posteriori* that the underlying assumptions are satisfied. Specifically, one should check the universality of the luminosity function, the convergence of the gravitational field, the effect of sampling ("shot") noise, and the reproducibility of density maps derived by different tracers. See Yahil (1986, 1988), Yahil, Strauss, and Davis (1990), Davis, Strauss, and Yahil (1990), and Babul and Postman (1990) for a discussion of these and other issues.

3. HOW TO COMPARE DENSITY AND VELOCITY SURVEYS

Once a density map has been created, one seeks to compare it with the observed velocity field. The next section describes the comparison with the smooth velocity field derived from the observed peculiar velocities using the *POTENT* technique. In this section I address general issues, which are applicable to any such comparison.

1. *Which Reference Frame For Peculiar Velocities?* The calculation of the gravitational field assumes that it is dominated by density inhomogeneities inside the volume surveyed, and inhomogeneities outside it may be ignored. In a multipole expansion of the external gravitational field, the leading term is the dipole, whose strength may be non-negligible (Vittorio and Juszkiewicz 1987; Juszkiewicz, Vittorio and Wyse 1990; Lahav, Kaiser, and Hoffman 1990). This dipole affects the comparison of the predicted velocity of the LG with the one deduced from the anisotropy of the CMB. The effect of this unknown dipole term can be removed, however, if peculiar velocities are evaluated in the LG frame, because the dipole velocity field is constant, and cancels out for all *relative* velocities. Quadrupole and higher terms due to the external field still result in systematic errors, of course, but these fall off faster with distance than the dipole field, and are generally negligible for distances less than half the survey limit. Note also that the transformation from redshift space to configuration space, Eq. (3), involves only velocities in the LG frame.

The comparison of the predicted velocities with the observed ones thus breaks down naturally into a prediction of the velocity of the LG relative to the CMB, and the velocities of other galaxies relative to the LG. The observations can also be divided in this fashion. The velocities of external galaxies are measured in the heliocentric frame, and are then converted either to the CMB or LG frame. The advantage of the CMB frame is the accuracy with which the solar velocity relative to it is known from measurements of the CMB dipole (Lubin and Villela 1986), while the solar velocity relative to the barycenter of the LG is uncertain by at least 50 km s^{-1} and perhaps 100 km s^{-1} (Yahil, Tammann, and Sandage 1977). When comparing with the density field, however, the advantage of the CMB frame is offset by the larger uncertainty in the predicted LG motion (cf., Yahil, Strauss, Davis, and Huchra 1990). I therefore concentrate on the velocity field in the LG frame.

2. *Compare Velocities or Densities?* The observed peculiar velocity and density fields can not be compared directly. Either the observed densities are converted to predicted velocities which are compared with the observed velocities, or, conversely, the observed velocities are converted to predicted densities which are compared with the observed densities. (In the former case the velocities should all be in the LG frame, for the reasons stated above.) There are advantages and disadvantages to both approaches.

The transformation from velocity to density requires spatial derivatives of the velocity field, a divergence in the linear approximation, and a determinant of the Jacobian in approximations based on that of Zel'dovich (1970; see also Nusser, Dekel, and Bertschinger 1990). Since the observed velocity data are noisy, this introduces large errors, except where the velocity field is averaged over a sufficient number of galaxies. This argues for a comparison of velocities.

If the noise in the observed peculiar velocity field can be overcome, however, there are several reasons for making a comparison of densities. First, the comparison is independent of reference frame. Secondly, it is a *local* comparison. As such, it is both insensitive to density structures outside the surveyed volume, and can serve to test the gravitational instability model itself. Moreover, density comparisons can also be made with optical surveys, which suffer from incomplete sky coverage, preventing a reliable prediction of the velocity field. (One is, of course, not totally exempt from computing the velocity field, albeit from a different dataset, because it is required for the transformation from redshift to configuration space.) A third advantage of a density comparison is the relative ease of computing nonlinear corrections to the transformation from velocity to density, while the nonlinear corrections in the reverse transformation are much more difficult (Yahil 1990; Nusser *et al.* 1990). Finally, errors in the transformation from redshift to configuration space can be detected by observing the relative displacement of structures in the density field.

3. *Zero Point Shift.* Irrespective of whether velocities or densities are compared, there is expected to be a zero-point shift between the compared fields. The density field is normalized within its survey volume. Peculiar velocities are calculated relative to a mean Hubble expansion of a set of galaxies, which typically do not sample the same volume. This introduces a possible zero point shift, which is a constant in density space, and mimics a Hubble expansion in velocity space.

4. COMPARISON WITH POTENT

Bertschinger and Dekel have developed a novel method for extracting a smoothed version of the peculiar velocity field from the data, known as *POTENT* (Bertschinger and Dekel 1989; Dekel, Bertschinger, and Faber 1990; Bertschinger, Dekel, Faber, and Dressler 1990;

the review by Bertschinger 1990, in this volume). In collaboration with them, this smoothed velocity field is used here to compare with the density field derived from the *IRAS* survey.

The simplest comparison is between density maps. Fig. 1 shows the *IRAS* and *POTENT* density maps in the Supergalactic Plane, smoothed by a 1200 km s^{-1} Gaussian. There is broad agreement between the maps, especially along the negative x-axis, in the direction of the "Great Attractor" in the Hydra-Centaurus region. In the opposite direction, the *IRAS* map shows a comparable density enhancement centered on Perseus-Pisces. The *POTENT* map is unreliable there, but some structure is seen where there are data in the Pisces direction.

Unfortunately, such visual comparisons of maps can be misleading, because the errors in each map are not obvious, except for the limits marked in the *POTENT* map. The rest of this section is therefore devoted to a quantitative comparison of the density and velocity fields, which takes full account of these errors. For both density and velocity, the comparison between *IRAS* and *POTENT* is made on a grid spaced by 500 km s^{-1} (in configuration space). *IRAS* densities, δ^I, are computed at the grid points by applying a Gaussian point-spread-function to each galaxy, whose width is independent of distance. As explained above, the "mass" per galaxy is proportional to the inverse of the selection function at the distance of the galaxy, and the survey extends to 8000 km s^{-1}. (See Yahil, Strauss, and Davis 1990 for further details.) *IRAS* velocities, $\mathbf{V}^I$, are computed from this smoothed density distribution using linear theory, Eq. (1), with $\lambda = 1$. *POTENT* densities, δ^P, and velocities, $\mathbf{V}^P$, are computed on the same grid with a Gaussian tensor-window-function whose width is 1200 km s^{-1}. The potent densities are also calculated assuming $\lambda = 1$.

For linear perturbations there is a simple relation between the two densities and two velocities:

$$\mathbf{V}^P = \lambda \mathbf{V}^I + a\mathbf{r} \tag{4a}$$

$$\delta^P = \lambda \delta^I - 3a \tag{4b}$$

where λ is the dynamical parameter sought, and a is the zero-point shift discussed in the last section. An estimate of λ is obtained by minimizing a weighted sum of residuals at the grid points:

$$\Lambda = \begin{cases} \sum_i w_i \left(\delta_i^P - \lambda\delta_i^I - 3a\right)^2 & \text{density} \\ \sum_i w_i \left(\mathbf{V}_i^P - \lambda\mathbf{V}_i^I + a\mathbf{r}\right)^2 & \text{velocity} \end{cases} \tag{5}$$

The weights are set equal to the inverse square of the estimated *POTENT* errors, based on Monte Carlo noise simulations for these data. The sum in Eq. (5) is further restricted by excluding grid points whose density error estimates $\Delta\delta^P > 0.2$, or which have fewer than 4 galaxies within 1500 km s^{-1}(cf., Fig. 1). Note that the quantity Λ in Eq. (5) should not be viewed as a χ^2, because the smoothed densities, or velocities, are correlated between adjacent grid points; the effective degrees of freedom are much smaller than the number of grid points.

The *IRAS* and *POTENT* smoothing processes both introduce additional errors, and the derived value of λ may be biased, as illustrated in the following figures. Fig. 2 shows
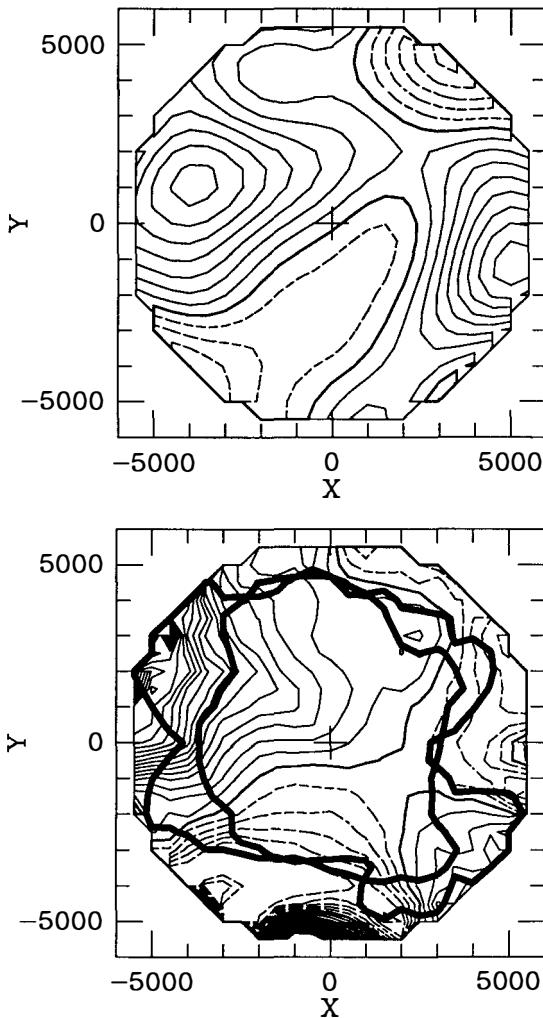
Fig. 1: Density fields in the Supergalactic Plane, smoothed by a 1200 km s^{-1} Gaussian: *IRAS* (top) and *POTENT* (bottom). Contours are spaced by 0.1 in δ, with positive ones solid, negative ones dashed, and the null contour, $\delta = 0$, thick solid. For *POTENT* two extra-thick contours mark limits, one of density error $\Delta\delta < 0.2$ (usually the more restrictive limit), and the other one of the existence of 4 data points within 1500 km s^{-1} from the grid point. The large Cross at the origin marks the position of the Local Group.

how well *POTENT* and *IRAS* match each other if there is adequate and noiseless sampling.
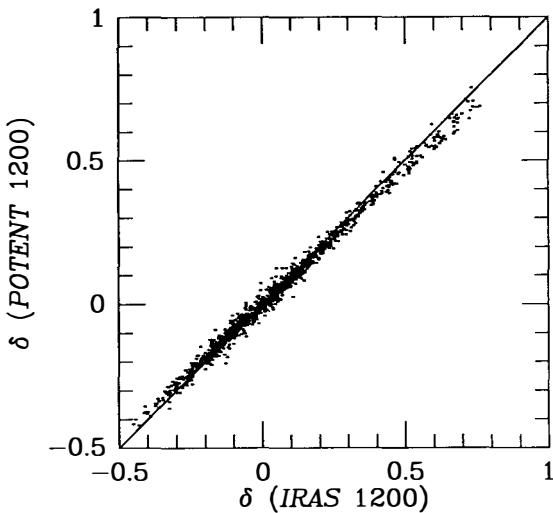
Fig. 2: Density derived by a linear version of *POTENT*, from the linear velocity field predicted by *IRAS* on the comparison grid itself, versus the *IRAS* density used to compute the velocity field.
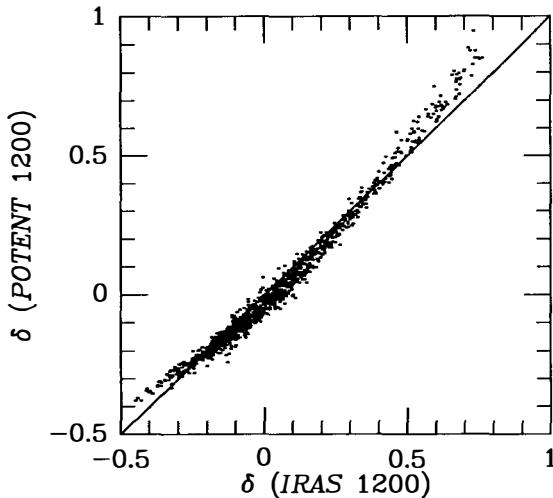

Fig. 3: Same as Fig. 2, but using a nonlinear version of *POTENT*.

In this case the velocities predicted by *IRAS* at the grid points themselves are used by a linear version of *POTENT*. The errors are mainly due to finite differencing in *POTENT*, and are negligible. Fig. 3 shows the same plot as Fig. 2, but with the standard nonlinear

version of *POTENT*. The scatter is comparable, but there is a noticeable curvature, due to incompatibility between the linear velocities predicted by *IRAS* and the nonlinear analysis performed by *POTENT*. As discussed in detail below, this is an important source of bias, which has not yet been fully calibrated.
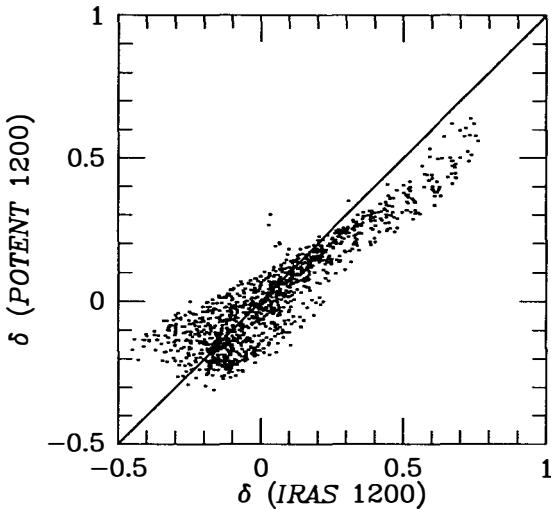

Fig. 4: Same as Fig. 3, but with predicted *IRAS* velocities provided only at the positions for which there exist observed peculiar velocities in the compendium circulated by D. Burstein.

Fig. 4 shows the same plot as Fig. 3, but with the predicted *IRAS* velocities provided only at the positions for which there exist observed peculiar velocities, as compiled by D. Burstein. The scatter is significantly larger, due to sparser sampling by *POTENT*. More importantly, a regression of δ^P on δ^I gives a slope significantly smaller than unity. In fact, it might be argued that the effective smoothing of *POTENT* densities is stronger than that given nominally by the tensor window function (Dekel *et al.* 1990), and the comparable *IRAS* smoothing should be larger. Fig. 5 is equivalent to Fig. 4 with a smoothing of 1400 km s^{-1}, and seems to fit better.

Finally, Fig. 6 shows the scatter diagram for the real galaxies, using the standard 1200 km s^{-1} tensor-window-function for *POTENT*, and 1400 km s^{-1} smoothing for *IRAS*. There is considerably more scatter, due to statistical uncertainties in the distances of the *POTENT* galaxies, and hence their deduced peculiar velocities.

The above figures raise several questions about the correspondence between *POTENT* and *IRAS*. How does the smoothing length in one scheme correspond to that in the other one? Are there any additional biases? What are the confidence limits on λ? These questions can not be answered analytically, but only by Monte Carlo simulations.
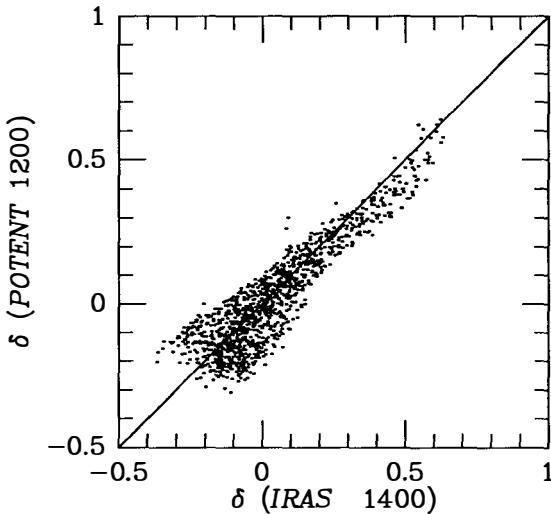
**Fig. 5:** Same as Fig. 4 with the *IRAS* data smoothed by 1400 km s⁻¹.
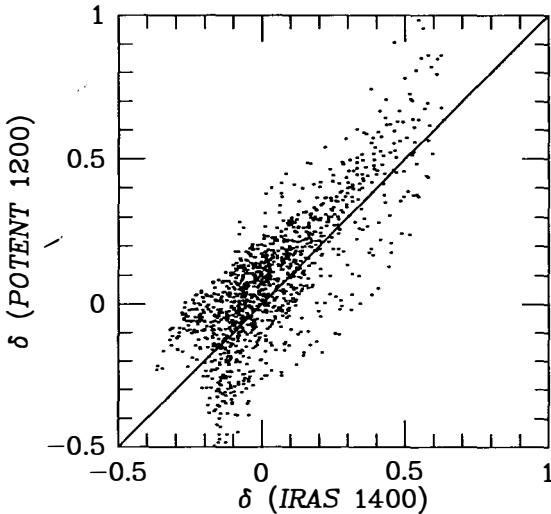
Fig. 6: *POTENT* versus *IRAS* densities for the real data.

In the Monte Carlo simulations used here, the *IRAS* density and velocity fields are taken as the null-hypothesis, and statistical noise is introduced into both fields as follows. The *IRAS* density field is varied by generating bootstrapped fields in which each galaxy is replaced by k galaxies, where k is a Poisson deviate with an expectation value of unity. This

replacement is only done in the final distribution in configuration space, assuming that the bulk of the *IRAS* error is due to sparse sampling. Errors induced by the transformation from redshift space to configuration space, as well as the addition of artificial "galaxies" in the 12% of the sky not surveyed by *IRAS*, are not taken into account. A full bootstrap of the original sample of observed galaxies, including all the transformations, is now underway.

Velocities predicted from linear theory, Eq. (1), are computed at the positions of the galaxies used by *POTENT* from the original (not bootstrapped) *IRAS* density distribution, using $\lambda = 1$. This provides *IRAS* estimates for the distances of these galaxies. The Monte Carlo simulations then scatter the distances using a Gaussian distribution whose standard deviation equals the distance uncertainty in the Tully-Fisher or $D_n - \sigma$ relation, as appropriate. These distances are inputted into *POTENT* as though they are the observed distances.

The Monte Carlo simulations are then run through the *IRAS* and *POTENT* procedures in order to generate for each densities and velocities on the grid, and a linear regression of the potent field on the *IRAS* field is performed by minimizing Λ in Eq. (5). If the procedure is unbiased, the derived value for the slope of the regression should be the input one, $\lambda = 1$. The degree to which the simulated λ's differ from unity measures the bias, and the scatter in their values provides confidence limits.

Table 1: Mean and Range of λ from 100 Monte Carlo Simulations

Field	*IRAS* Smoothing	Mean	5th percentile	95th percentile
Density	1200 km s^{-1}	0.78	0.56	0.97
Density	1400 km s^{-1}	0.99	0.62	1.24
Density	1500 km s^{-1}	1.10	0.78	1.41
Velocity	1200 km s^{-1}	0.96	0.80	1.15
Velocity	1300 km s^{-1}	0.97	0.77	1.18

The results of 100 Monte Carlo simulations are listed in Table 1. The derived values of λ are seen to be strongly biased for the density comparisons, and mildly so for the velocity comparisons. In order to correct for the bias, the values derived for the real data are scaled by dividing by the mean of the Monte Carlo simulations. The results are shown in Table 2. The confidence range is taken to be proportional to the one in the Monte Carlo simulations.

Two conclusions emerge from Table 2. First, the bias-corrected values of λ do not depend on the *IRAS* smoothing used; hence the bias seems to be adequately removed by the Monte Carlo correction. More disturbing is the inconsistency between the values of λ derived by comparing densities and by comparing velocities. This discrepancy is not yet fully understood, but part of the explanation undoubtedly involves nonlinear corrections. Recall that *IRAS* velocities are computed from linear theory, Eq. (1), while Nature and *POTENT* include nonlinear effects. For a given density structure, linear theory overestimates peculiar velocities (Yahil 1985; Villumsen and Davis 1986; Regős and Geller 1989; Lightman and Schechter 1990; Yahil 1990). When the *IRAS* predicted velocities are compared with

Table 2: Observed and Bias-Corrected λ

Field	IRAS Smoothing	Observed λ	Bias-Corrected λ
Density	1200 km s^{-1}	0.79	1.01 (+0.26 –0.27)*
Density	1400 km s^{-1}	0.99	1.00 (+0.25 –0.38)
Density	1500 km s^{-1}	1.10	0.99 (+0.26 –0.30)
Velocity	1200 km s^{-1}	0.72	0.74 (+0.15 –0.12)
Velocity	1300 km s^{-1}	0.72	0.75 (+0.16 –0.15)

* 5%–95% confidence limit.

the observations, therefore, λ is underestimated. At first glance it might appear that this argument should not apply to the comparisons of the density fields, since the linear IRAS velocities are not used, except in the transformation from redshift to configuration space. They are used, however, in the Monte Carlo simulations which calibrate the bias correction. It is not easy to estimate this calibration error, or even its sign, because it can affect both λ and the zero-shift parameter a in Eq. (4). Nonlinear IRAS velocity estimates (Yahil 1990) may help in this regard.

Until the discrepancy between the values of λ derived from density and from velocity is fully resolved, it may be safer to conclude that the true value of λ probably lies between the values obtained by the two methods, say, $\lambda = 0.75 - 1.15$, to allow for both statistical and systematic error margins. For unbiased galaxy formation this range corresponds to $\Omega = 0.6 - 1.25$.

5. OTHER DETERMINATIONS OF Ω

A number of estimates of Ω, or, equivalently, the bias parameter b, have been discussed over the years (e.g., the review by Yahil 1987). Most of the estimates have been significantly lower than the range $\Omega = 0.6 - 1.25$ determined in the last section. Are these real discrepancies, or do either the new or the older values need to be re-examined? The following is a brief compendium of previous, lower, Ω estimates, and problems associated with them.

1. *Big-Bang Nucleosynthesis.* The comparison of the observed abundances of the light elements with the predictions of big-bang nucleosynthesis seem to provide a firm limit on the *baryonic* content of the universe, $0.0097 < \Omega_B h^2 T_{2.75}^{-3} < 0.016$ (Olive *et al.* 1990; Schramm 1990, in this volume). In the standard interpretation, the difference between this value and the total Ω is ascribed to dark matter; other possibilities include primordial black holes, nonstandard nucleosynthesis, or remaining systematic errors in the abundances of the light elements.

2. *Halos of Galaxies.* If the typical masses of a galaxies were known, as well as their number density in the universe, one could deduce the mass density in galaxies. Unfortunately, the rotation curves of galaxies are flat to the farthest galactocentric radii observed, and show no

Keplerian falloff (e.g., Rubin *et al.* 1980, 1982, 1985; van Albada *et al.* 1985; Carignan and Freeman 1985), so we do not know their total masses. Nor do binary galaxies provide any clue (e.g., Rivolo and Yahil 1981; White *et al.* 1983). We also do not know how to weigh the dwarf galaxies, which are very numerous. A lower limit $\Omega \gtrsim 0.1$ is obtained by assuming that a typical bright galaxy, $L \sim L_*$ in the language of Schechter (1976), extends to at least 100 kpc, but this estimate is very uncertain.

3. *Clusters of Galaxies.* Recent determinations of the mass distribution in clusters of galaxies, taking into account both the X-ray and optical data (Hughes 1989; Fabricant, Kent, and Kurtz 1989) show the mass-to-light ratio to be $M/L_B = 200 - 400h$ $M_\odot/L_\odot$. Combining this with the mean luminosity density (emissivity) of the universe, $\mathcal{L} = 1.93^{+0.8}_{-0.6} \times 10^8 h^{-1}$ Mpc^{-3} (Efstathiou, Ellis, and Peterson 1988), one obtains $\Omega_G = 0.1 - 0.3$. This estimate is compromised, however, by the change in morphological abundances in rich clusters, which are dominated by elliptical and lenticular galaxies, whereas the bulk of the galaxies contributing to the luminosity density of the universe are spirals (Dressler 1980).

Clusters of galaxies can also provide an estimate of the ratio of baryonic to total (dynamical) mass. Hughes (1989) determined the gas mass and total mass in the inner $0.5h^{-1}$ Mpc of the Coma cluster, as well as the blue luminosity of the galaxies within that radius. (This radius, a third of an Abell radius, was chosen because both the gas mass and the total mass can be determined more securely, and the derived values are relatively insensitive to assumptions.) Taking a mass-to-light ratio $M/L = 10h$ $M_\odot/L_\odot$ for the stellar matter in elliptical galaxies (Lauer 1985; see the discussion in Binney and Tremaine 1987), the results of Hughes constrain the ratio of baryonic to total mass to be in excess of $0.035h^{-3/2} + 0.025$. The first term in this lower limit is the contribution of the gas, and the second term that of the stars in the galaxies. If this represents a lower limit on the general baryonic fraction in the universe, then it may be difficult to reconcile the results of big-bang nucleosynthesis with a total $\Omega = 1$. A possible explanation might be the existence of massive cooling flows, but the amount of gas flowing in would need to be very large, and the present density profile of the gas in Coma, with the well-defined core radius seen in the X-rays, argues against this.

4. *"Cosmic Virial Theorem".* This method estimates Ω by comparing the r.m.s. pairwise velocity differences of neighboring galaxies with a statistical calculation of their relative accelerations, using the three-point galaxy correlation function (Peebles 1976*b*). The method is typically applied to galaxies whose separations, projected on the plane of the sky, is less than $1h^{-1}$ Mpc. There are technical difficulties with the three-point correlation function (Rivolo and Yahil 1981; Davis and Peebles 1983), but, more importantly, dynamical friction can bias the estimate of Ω. When the halos of galaxies come in contact with one another they merge on a dynamical timescale, which is typically much smaller than the age of the universe. What we observe as galaxies, namely the high density luminous centers, are not immediately disrupted, but dynamical friction causes them to spiral into the bottom of the potential well of the mass distribution. The observed galaxies are therefore more concentrated than the mass distribution (Barnes 1983, 1984; Evrard 1986, 1987). (Note that this bias applies only

to highly nonlinear perturbations, and is not related to biased galaxy formation in the linear regime.) It has recently been claimed that dynamical friction also results in a slowdown of velocities, thus further decreasing mass estimates (Carlberg and Couchman 1989; Carlberg, Couchman, and Thomas 1990).

5. **N-body Simulations.** N-body simulations with $\Omega = 1$ systematically generate "hot" models, in which the velocity dispersion of galaxies is much higher than observed. Their Mach number is too high, in the language of Ostriker and Suto (1990). In recent years attempts have been made to overcome this problem by introducing "biasing" to the interpretation of the simulations (e.g., Davis *et al.* 1985). I suspect that the fault lies in the resolution of the simulations, in which the mass per particle is typically a galactic mass. The rapidly improving computer technology now allows simulations in which there are many particles per galaxy, augmented by hydrodynamics of the gas, using smooth particle hydro (SPH) methods. These should provide a more realistic simulation of the interaction of galaxies, including effects of merges, dynamical friction and so on.

6. **Virgocentric Infall.** This flow model in the Local Supercluster gave some of the strongest early arguments for a low value of $\Omega = 0.1 - 0.2$ (e.g., the review by Yahil 1985). We now know that the model is too naive. The problem is not with the spherical approximation to the potential of the density enhancement around Virgo, which is valid, but the neglect of other sources of gravity. First, the tidal field of neighboring superclusters is important (Lilje, Yahil, and Jones 1986), although fortuitously, owing to the alignment of these superclusters in the vicinity of the Local Supercluster they affect the velocity of the LG relative to the Virgo cluster very little. The real culprit is the local void in the North Supergalactic Hemisphere (Yahil, Sandage, and Tammann 1980; Tully and Fisher 1987), which did not receive enough attention in the early years.

7. **Galaxy Dipole Anisotropies versus the CMB Dipole.** The comparison of dipole anisotropies in the distribution of galaxies with the dipole anisotropy of the CMB has received considerable attention in recent years. Galaxy samples have been selected both in the infrared (Yahil, Walker, and Rowan-Robertson 1986; Meiksin and Davis 1986; Strauss and Davis 1988a,b; Yahil 1988; Villumsen and Strauss 1987; Strauss 1989) and optical (Lahav 1987; Dressler 1988; Lahav, Rowan-Robinson, and Lynden-Bell 1988). Ω estimates are typically lower for the optical surveys than for the infrared ones. Unfortunately, the dipole comparison is compromised by possible largescale power beyond the range surveyed by the galaxies (Vittorio and Juszkiewicz 1987; Juszkiewicz, Vittorio and Wyse 1990; Lahav, Kaiser, and Hoffman 1990), as well as by smallscale smoothing (Yahil, Strauss, Davis, and Huchra 1990). Irrespective of that, the optical surveys are subject to severe problems of sky coverage, Galactic absorption, lack of redshifts, and imprecise selection by magnitude and/or angular diameter.

8. **Morphological Segregation.** *IRAS* galaxies, which are spirals, seriously underestimate the density in rich clusters of galaxies. Morphological segregation, however, begins only at

densities that are a factor of a hundred or more in excess of the mean density of the universe (Dressler 1980; Yahil, Sandage, and Tammann 1980; Postman and Geller 1984). This is confirmed by direct comparisons of *IRAS* and optical densities (Yahil 1986; Strauss 1989; Babul and Postman 1990), which show them to trace each other well, except in a few rich clusters. Most of the galaxies are not in these high density peaks, and the effect of this bias should not be too large. Attempts to correct for this bias by double counting galaxies in clusters (Strauss and Davis 1988*b*), and also around them Strauss (1989), indeed yield lower estimates of Ω, but in some cases the double counting may have been excessive.

9. *Biasing.* The really nagging question, which can only be answered dynamically, is whether any set of galaxies really trace the mass. As argued in the introduction, an important test is the *universality* of the relation between the velocity and density fields, predicted by gravitational instability theory. In order to test this universality, it is necessary to establish the relation in *several* density structures. The comparison performed in this paper, with smoothing in excess of 1000 km s^{-1}, is concentrated in the single mass conglomerate in the "Great Attractor". Other density structures to be probed include the Local Supercluster, with higher resolution, and the Perseus-Pisces complex, as more peculiar velocities become available in that region.

ACKNOWLEDGEMENTS

REFERENCES

Aaronson, M. *et al.* 1989, *Astrophys. J.*, **338**, 654.
Babul, A., and Postman, M. 1990, *Astrophys. J.*, in press.
Barnes, J. 1983, *Mon. Not. Roy. Astr. Soc.*, **203**, 223.
Barnes, J. 1984, *Mon. Not. Roy. Astr. Soc.*, **208**, 885.
Bertschinger, E. 1990, this volume.
Bertschinger, E., and Dekel, A. 1989, *Astrophys. J. (Letters)*, **336**, L5.
Bertschinger, E., Dekel, A., Faber, S. M., and Dressler, A. 1990, *Astrophys. J.*, in press.
Binney, J., and Tremaine, S. 1987, *Galactic Dynamics* (Princeton: Princeton University Press).
Carlberg, R. G., and Couchman, H. M. P. 1989, *Astrophys. J.*, **340**, 47.
Carlberg, R. G., Couchman, H. M. P., and Thomas, P. A. 1990, *Astrophys. J. (Letters)*, , in press.
Carignan, C., and Freeman, K. C. 1985, *Astrophys. J.*, **294**, 494.
da Costa, L. N., Pellegrini, P., Sargent, W., Tonry, J., Davis, M., Meiksin, A., Latham D., Menzies, J., and Coulson, I. 1988, *Astrophys. J.*, **327**, 544.

Davis, M., Efstathiou, G., Frenk, C. S., and White, S. D. M. 1985, *Astrophys. J.*, **292**, 371.

Davis, M., and Huchra, J. 1982, *Astrophys. J.*, **254**, 437.

Davis, M., and Peebles, P. J. E. 1983, *Ann. Rev. Astr. Ap.*, **21**, 109.

Davis, M., Strauss, M. A., and Yahil, A. 1990, *Astrophys. J.*, submitted.

Dekel, A., Bertschinger, E., and Faber, S. M. 1990, *Astrophys. J.*, in press.

Dekel, A., and Rees, M. J. 1987, *Nature*, **326**, 455.

Dressler, A. 1980, *Astrophys. J.*, **236**, 351.

Dressler, A. 1988, *Astrophys. J.*, **329**, 519.

Efstathiou, G., Ellis, R. S., and Peterson, B. A. 1988, *Mon. Not. Roy. Astr. Soc.*, **232**, 431.

Evrard, A. E. 1986, *Astrophys. J.*, **310**, 1.

Evrard, A. E. 1987, *Astrophys. J.*, **316**, 36.

Fabricant, D. G., Kent, S. M., and Kurtz, M. J. 1989, *Astrophys. J.*, **336**, 77.

Hughes, J. P. 1989, *Astrophys. J.*, **337**, 21.

Juszkiewicz, R., Vittorio, N., and Wyse, R. F. G. 1990, *Astrophys. J.*, **349**, 408.

Lahav, O. 1987, *Mon. Not. Roy. Astr. Soc.*, **225**, 213.

Lahav, O., Kaiser, N., and Hoffman, Y. 1990, *Astrophys. J.*, **352**, 448.

Lahav, O., Rowan-Robinson, M., and Lynden-Bell, D. 1988, *Mon. Not. Roy. Astr. Soc.*, **234**, 677.

Lauer, T. R. 1985, *Astrophys. J.*, **292**, 104.

Lightman, A., and Schechter, P. 1990, *Astrophys. J.*, in press.

Lilje, P. B., Yahil, A., and Jones, B. J. T. 1986, *Astrophys. J.*, **307**, 91.

Lubin, P. M., and Villela, T. 1986, in *Galaxy Distances and Deviations from Universal Expansion*, ed. B. F. Madore and R. B. Tully (Dordrecht: Reidel), p. 169.

Lynden-Bell, D., Faber, S. M., Burstein, D., Davies, R. L., Dressler, A., Terlevich, R. J., and Wegner, G. 1988, *Astrophys. J.*, **326**, 19.

Meiksin, A., and Davis, M. 1986, *Astr. J.*, **91**, 191.

Nusser, A., Dekel, A., and Bertschinger, E. 1990, *Astrophys. J.*, submitted.

Olive, K. A., Schramm, D. N., Steigman, G., and Walker, T. P. 1990, *Phys. Lett.*, **B236**, 454.

Ostriker, J. P., and Suto, Y. 1990, *Astrophys. J.*, **348**, 378.

Peebles, P. J. E. 1976a, *Astrophys. J.*, **205**, 318.

Peebles, P. J. E. 1976b, *Astrophys. J. (Letters)*, **205**, L109.

Postman, M., and Geller, M. 1984, *Astrophys. J.*, **281**, 95.

Regős, E., and Geller, M. J. 1989, *Astr. J.*, **98**, 755.

Rivolo, A. R., and Yahil, A. 1981, *Astrophys. J.*, **251**, 477.

Rubin, V. C., Burstein, D., Ford, W. K., and Thonnard, N. 1985, *Astrophys. J.*, **289**, 81.

Rubin, V. C., Ford, W. K., and Thonnard, N. 1980, *Astrophys. J.*, **238**, 471.

Rubin, V. C., Ford, W. K., Thonnard, N., and Burstein, D. 1982, *Astrophys. J.*, **261**, 439.

Schechter, P. L. 1976, *Astrophys. J.*, **203**, 297.

Schramm, D. N. 1990, this volume.

Strauss, M. A. 1989, Ph. D. thesis, University of California, Berkeley.

Strauss, M. A., and Davis, M. 1988a, in Proc. IAU Symp. # 130, *Large Scale Structures of the Universe*, ed. J. Audouze, M. -C. Pelletan, and A. Szalay (Dordrecht: Reidel), p. 191.

Strauss, M. A., and Davis, M. 1988b, in *Large Scale Motions in the Universe: A Vatican Study Week*, edited by V. C. Rubin and G. V. Coyne, (Princeton: Princeton University Press), p. 256.

Strauss, M. A., Davis, M., Yahil, A., and Huchra, J. P. 1990, *Astrophys. J.*, in press.

Tully, R. B., and Fisher, J. R. 1987, *Nearby Galaxies Atlas* (Cambridge: Cambridge University Press).

van Albada, T. S., Bahcall, J. N., Begeman, K., and Sancisi, R. 1985, *Astrophys. J.*, **295**, 305.

Villumsen, J. V., and Davis, M. 1986, *Astrophys. J.*, **308**, 499.

Villumsen, J. V., and Strauss, M. A. 1987, *Astrophys. J.*, **322**, 37.

Vittorio, N., and Juszkiewicz, R. 1987, in *Nearly Normal Galaxies*, ed. S. M. Faber (New York: Springer), p. 451.

White, S. D. M., Huchra, J. P., Latham, D., and Davis, M. 1983, *Mon. Not. Roy. Astr. Soc.*, **203**, 701.

Yahil, A. 1985, in *The Virgo Cluster of Galaxies*, eds. O. G. Richter and B. Binggeli (Garching: European Southern Observatory).

Yahil, A. 1986, in *Galaxy Distances and Deviations from the Hubble Expansion*, ed. B. Madore, (Dordrecht: Reidel), pp. 151–158.

Yahil, A. 1987, in *Nearly Normal Galaxies: From the Planck Time to the Present*, ed. S. M. Faber (Berlin: Springer), pp. 332–342.

Yahil, A. 1988, in *Large Scale Motions in the Universe: A Vatican Study Week*, edited by V. C. Rubin and G. V. Coyne, S. J. (Princeton: Princeton University Press), p. 219.

Yahil, A. 1990, in preparation.

Yahil, A., Sandage, A., and Tammann, G. A. 1980, *Astrophys. J.*, **242**, 448.

Yahil, A., Strauss, M. A., and Davis, M. 1990, *Astrophys. J.*, submitted.

Yahil, A., Strauss, M. A., Davis, M., and Huchra, J. P. 1990, *Astrophys. J.*, in preparation.

Yahil, A., Tammann, G. A., and Sandage, A. 1977, *Astrophys. J.*, **217**, 903.

Yahil, A., Walker, D., and Rowan-Robinson, M. 1986, *Astrophys. J. (Letters)*, **301**, L1.

Zel'dovich, Ya. B. 1970, *Astr. Astrophys.*, **5**, 84.

LIST OF PARTICIPANTS

ALIMI Jean-Michel
Observatoire de Paris
DARC
F- 92195 MEUDON Cedex
FRANCE

AUBOURG Eric
Institut d'Astrophysique de Paris
98 bis Boulevard Arago
F- 75014 PARIS
FRANCE

BAGIEU Guy
Université de Grenoble
Inst. des Sciences Nucléaires
F- 38026 GRENOBLE Cedex
FRANCE

BELLI Pierluigi
Univ. di Roma "Tor Vergata"
Dipart. di Fisica
I- 00173 ROMA
ITALY

BERTSCHINGER Edmund
M. I. T.
Dept of Physics
CAMBRIDGE MA 02139
USA

BLANCHARD Alain
Observatoire de Paris-Meudon
Service d'Astrophysique
F- 92195 MEUDON PRINCIPAL Cedex
FRANCE

BOISSEAU Bruno
Université de Tours
Faculté des Sciences
F- 37200 TOURS
FRANCE

BONOMETTO Silvio
Universita di Perugia
Dipt di Fisica
I- 06100 PERUGIA
ITALY

BORGANI Stefano
Scuola Internazionale Superiore
di Studi Avanzati (SISSA)
I- 34100 TRIESTE
ITALY

BOUQUET Alain
Université Paris VII
LPTHE - Tour 24 - 5ème Etage
F- 75251 PARIS
FRANCE

BURKE Bernard
M. I. T.
Dept of Physics
CAMBRIDGE MA 02139
USA

CARR Bernard
University of London
Queen Mary College - Applied Math. Dept.
LONDON E1 4NS
UNITED KINGDOM

CARTER Brandon	Observatoire de Paris-Meudon Service d'Astrophysique F- 92195 MEUDON PRINCIPAL Cedex FRANCE
COMBES Françoise	Observatoire de Paris-Meuon Radioastronomie F- 92195 MEUDON PRINCIPAL Cedex FRANCE
DE RUJULA Alvaro	CERN TH Division CH- 1211 GENEVE 23 SWITZERLAND
DOLGOV Alexandre	ITEP Inst. for Theor. and Exp. Physics 117259 MOSCOW U S S R
GIRAUD-HERAUD Yannick	Collège de France Lab. de Physique Corpusculaire F- 75231 PARIS Cedex 05 FRANCE
GODLOWSKI Wlodzimiorz	Uniwersytetu Jagiellonskiego Observatory Astronomiczne PL- 30-244 KRAKOW POLAND
GONZALEZ-MESTRES Luis	LAPP Lab. de Physique des Particules F- 74941 ANNECY LE VIEUX Cedex FRANCE
GUIDERDONI Bruno	Institut d'Astrophysique de Paris 98 bis Boulevard Arago F- 75014 PARIS FRANCE
JETZER Philippe	CERN TH Division CH- 1211 GENEVE 23 SWITZERLAND
KAPLAN Jean	Université Paris VII LPTHE - Tour 24 - 5ème Etage F- 75251 PARIS FRANCE
KHLOPOV Maxim	Keldysh Institute of Applied Mathematics Miusskaya Pl. 4 125047 MOSCOW U S S R
KOOPMAN Rob	University of Amsterdam Astronomical Institute NL- 1018 WB AMSTERDAM NETHERLANDS

KRAUSS Lawrence

Yale University
Sloane Lab. - Dept of Physics
NEW HAVEN   CT 06520
USA

LESQUOY Eric

CEN Saclay
DPhPE
F- 91191 GIF-SUR-YVETTE Cedex
FRANCE

LINDE Andrei

CERN
TH Division
CH- 1211 GENEVE 23
SWITZERLAND

LONGARETTI Pierre-Yves

Observatoire Midi Pyrennées
14, avenue E. Belin
F- 31400 TOULOUSE
FRANCE

MADSEN Jes

Astronomisk Institute
Universitetsparken Langelandsgade
DK- 8000 AARHUS C
DENMARK

MARTIN de VOLNAY François

Université Paris VII
LPTHE - Tour 24 - 5ème Etage
F- 75251 PARIS
FRANCE

MATARRESE Sabino

Universita di Padova
Istituto di Fisica - Galilei
I- 35191 PADOVA
ITALY

MOLLERACH Silvia

Scuola Internazionale Superiore
di Studi Avanzati (SISSA)
I- 34100 TRIESTE
ITALY

MONIEZ Marc

Université Paris Sud
LAL
F- 914O5 ORSAY Cedex
FRANCE

MOUTARDE Fabien

Observatoire de Paris Meudon
DAEC
F- 92195 MEUDON PRINCIPAL Cedex
FRANCE

NOTTALE Laurent

Observatoire de Paris Meudon
DAEC
F- 92195 MEUDON PRINCIPAL Cedex
FRANCE

OLIVE Keith

University of Minnesota
Dept of Physics and Astronomy
MINNEAPOLIS MN 55455
USA

PEACOK John

University of Edinburgh
Royal Observatory
EDINBURGH EH9 3HJ
UNITED KINGDOM

PETER Patrick

Observatoire de Paris Meudon
DAEC
F- 92195 MEUDON PRINCIPAL Cedex
FRANCE

PETROSIAN Vahé

University of Stanford
Dept of Physics
STANFORD CA 94305
USA

PIRAN Tsvi

Hebrew University
Recah Institut of Phsyics
91904 JERUSALEM
ISRAEL

PROVENZALE Antonello

CNR
Istituto di Cosmo-Geofisica
I- 10125 TORINO
ITALY

RAFFELT Georg

Max Planck Institut
Physik und Astrophysik
D- 8000 MUNCHEN 40
FEDERAL REP. OF GERMANY

SALATI Pierre

CERN
TH Division
CH- 1211 GENEVE 23
SWITZERLAND

SANCHEZ Norma

Observatoire de Meudon
DMIRM
F- 92195 MEUDON PRINCIPAL Cedex
FRANCE

SCAFI Marta

Univ. di Roma "Tor Vergata"
Dipart. di Fisica
I- 00173 ROMA
ITALY

SCHRAMM David

University of Chicago
Enrico Fermi Inst. - Dept of Physics
CHICAGO IL 60637
USA

SEGUI SANTONJA Antonio José

Universidad de Zaragoza
Depto de Fisica Teorica
E- 9 ZARAGOZA
SPAIN

SELIPSKY Stephen

University of Stanford
Dept of Physics
STANFORD CA 94305
USA

SIROUSSE ZIA Haydeh

Institut Henri Poincaré
Lab. de Physique Théorique
F- 75231 PARIS Cedex
FRANCE

TAO Charling

CEN Saclay
DPhPE
F- 91191 GIF-SUR-YVETTE Cedex
FRANCE

YAHIL Amos

State University of New York (SUNY)
Dept of Physics
STONY BROOK NY 11794
USA