Fermi National Accelerator Laboratory

FERMILAB-Pub-86/119-A
August 1986

PRIMEVAL QSOs

A. CAVALIERE[1] and A.S. SZALAY[2,3]

1. Astrofisica, Dip. Fisica II Universita di Roma
2. NASA/Fermilab Astrophysics Group
3. Eötvös University, Budapest

ABSTRACT. Motivated by preliminary data suggesting a decreasing surface density of QSOs beyond z ~ 2, we discuss how the turn-on time scales of active galactic nuclei may be related to the era of galaxy formation, and derive predictions for the z-behaviour of the number density of both classes of objects.

subject headings: cosmology - quasars - galaxies: formation.

Ap.J. in press

I. INTRODUCTION

QSOs as a population are known to increase steeply in apparent number at a given luminosity, as they are traced farther back into our past light cone (cf Schmidt and Green 1983, Setti and Zamorani 1984, Koo 1985).

A "luminosity evolution" of the population (which implies an average dimming or fading of all objects in cosmic time from a bright initial stage, with the object number *conserved*) complies with the current data body: the counts at bright and faint magnitudes, the redshift distributions up to $z \approx 2.2$, and the constraint of producing less than the full intensity of the keV X-ray background (primarily from sources of Seyfert-type luminosity). In this picture a moderate increase in luminosity toward high z: $L(z)/L(O) \sim 10^2$, entails a much larger increase in the apparent number $[L(z)/L(O)]^{\gamma-1}$ if the luminosity function $N(L,z)dL \propto L^{-\gamma}dL$ is steep (slope $\gamma \rightarrow 3.5$, Schmidt and Green 1983). The timescale required for the average object's fading is $t_f \approx 3$ Gyr (H_O = 50 km/s Mpc), to zeroth order constant and uniform at high luminosities ($L > 10^{45}$ erg/s). If in a next approximation t_f lengthens somewhat toward lower L, the XRB production is safely minimized, and $N(L,z)$ of the distant QSOs goes smoothly into the local $N(L,z \sim O)$ of Active Galactic Nuclei of Seyfert 1 type with its distinctly bent faint end, still conserving numbers (Cavaliere, Giallongo and Vagnetti 1985).

However, as noted also by all the above authors, new events must set in at and beyond $z \approx 2$: this is strongly indicated by the remarkable dearth of higher z objects, especially at faint magnitudes $B \gtrsim 20$ corresponding to medium and low luminosities.

II. THE PROBLEM

The specific observational evidence includes the deep surveys by Osmer (1982), Schneider, Schmidt and Gunn (1984), and Koo, Kron and Cudworth (1986). Koo (1985) finds, along with a flattening of $N(L,z)$ at its faint end, a strong evolution in L out to $z \approx 2$ but a possible "anti-evolution" in number for $z > 2$. From objective prism searches in larger fields Hazard and McMahon (1985) began to quantify the rate of decline for the more persistent, intrinsically bright objects (also above the survey's magnitude limit); they estimate a ~ 1/10 fall from $z \sim 2$ to 3, and possibly a further decline by some 1/3 from $z \sim 3$ to 3.6 (cf Fig 1). The apparent breakdown of conservation over such a short time interval, $t \approx 1$ Gyr $< H^{-1}(2)$ for $\Omega_0 = 0.2$ to 1 ($\lambda_0 = 0$), poses intriguing problems.

It may be argued that only the object visibility is affected, most likely by dusty intervening galaxies (cf discussions by Bechtold et al 1984, Ostriker and Heisler 1984, Netzer 1985). But then time lags and statistical spreads in dust production, plus the variance in galaxy cross sections along the line of sight, ought to smear out the obscured era and to cause much precursory reddening. In any case, medium and low *L* objects if still extensively present beyond z~2 would overproduce the XRB (cf Setti and Zamorani 1984). Actually, a stronger constraint is set by the *flat and smooth* spectral shape of the XRB at 2 - 20 keV, which would be spoiled beyond repair even by a › 1/3 summed contribution from objects with Seyfert-like or steeper X-ray spectra extending into the ~10 keV range (Cavaliere et al 1981, De Zotti et al 1982; cf subsequent data by Elvis, Wilkes and Tananbaum 1985).

In a similar vein, systematic spectral evolution in the early activity stage, including heavy intrinsic extinction, may well elude canonical selection and identification criteria in the optical band (Hazard et al 1984, Cavaliere Giallongo and Vagnetti 1985 and references therein). But once again, the (redshifted) contribution to the XRB is bound to be excessive if mere filtering of a canonical primary activity were all that happened from *z* ~2 to 4. If, on the other hand, spectral evolution occurs on very short scales such as provided by sweeping of lingering circumnuclear matter at $v \rightarrow 10^2 - 10^3$ km/s, then the problem of the decline time scale is simply deferred to the rate of turn-on for the primary activity.

In sum, the optical data strongly suggest, and the XRB constraints independently require, that any net increase of all objects (except possibly the ultrabright ones) should stop and reverse soon beyond *z*~ 2. A similar trend is shared by the radio loud QSRs included in the analysis

Peacock 1985 (see also Condon 1984), further relievieng worries that the whole of the decline might be due to selection effects specific to the optical band.

One may then consider a relatively weak but synchronous lunimosity decrease for $z>2$, amplified by the steep slope of $N(L,z)$ in a "luminosity anti-evolution" (a brightening up in cosmic time) that still conserves numbers. However, this solution by itself would only partially alleviate the timescale problem; it would affect more the brighter than the weaker objects presumably located on the flatter section of $N(L,z>2)$, an outcome at variance with the trend of the current upper limits at faint magnitudes. Certainly, actual detections of some such weaker objects are much wanted to separate the contributions of a luminosity and of a number increase for $z \gtrsim 2$ (Koo 1985).

But in a broader view such a brightening would constitute only a part of the whole turn on process, implying an even earlier, and still fast, object production. The aim of this paper is to focus on true *number increase* of QSOs in the range $z \sim 3$ to 2, and to discuss under what conditions a sharp z-dependence can be understood in terms of the physical processes governing their formation.

III. A TIME SCALE FOR TURN ON

If number conservation breaks down at $z>2$, the continuity equation governing the evolution of $N(L,t)$ must include a source term $S(L,t)$

$$\frac{\partial N}{\partial t} + \frac{\partial}{\partial L}(\dot{L}N) = S(L,t) \qquad (1)$$

The nature of the primary energy source is reflected, after some averaging, in the term $\dot{L}$. A strong candidate being accretion onto a massive black hole hosted in a galactic nucleus (see Rees 1984), the dimming time scale $t_f = L/|\dot{L}| \sim 3$ Gyr well under H_o^{-1} suggests the tapping of mass supplies either localized within ~1 pc of the hole, or related to longer range events acting over times $< H^{-1}(z)$.

In fact, mass inflows to feed $L \sim 10^{43}$ to 10^{45} erg/s may be provided by recurrent interactions of the host with companion galaxies, for which there is much (though not systematic) evidence mostly associated with objects of low or intermediate powers and redshifts (Dahari 1984, De Robertis 1985, Heckman et al 1984, Keel et al 1985). These interactions are likely to sustain the activity by refuelling an already formed hole; they may act through perturbations of stellar orbits from a much larger volume (Norman and Silk 1983), or by triggering cannibalism of satellite galaxies by the host (Gaskell 1985). For z -> 0 these interactions would rapidly weaken in a hierarchical clustering scenario (Roos 1981), since within the progressively larger associations of galaxies the member density thins down while the velocity dispersion grows.

But toward higher z, when $H^{-1}(z) \lesssim t_f$ holds, a description of the fuelling process in terms of weak, recurrent interactions no longer applies. Indeed, in order to reach the high luminosities ($L > 10^{46}$ erg/s) well represented among distant QSOs, the detailed models (see McMillan, Lightman and Cohn 1981, Duncan and Shapiro 1983, Shapiro and Teukolsky 1985) have assumed unusual concentrations of mass, 10^8 to 10^9 $M_\odot$ densely packed within ~1 pc. This is more likely to be material left over from the very formation process, out of which the holes grew substantially in mass and output over timescales $< t_f$: here the models should link directly with formation events of the AGNs.

The range of z involved and the mass density required suggest that such deep collapses of large masses should be related in turn to the overall collapse and settling of protogalaxies, specifically, to the formation of well defined gravitational centers. But there are uncertainties concerning the dynamical details: given proper initial conditions, still a number of different routes may lead to the formation of a hole of $M > 10^6 M_\odot$ (see Begelman and Rees 1978), and the associated time scale t_b varies accordingly. Definite values are provided by straight infall with $t_b > t_{dyn}$, or by continuous accretion at the formal Eddington rate which takes $t_b \lesssim 0.4$ Gyr * several exponentiations; but reduced efficiency or normal amounts of specific angular momentum may place t_b anywhere in the range ~ 10^{-2} to several Gyr (cf Rees 1984). It may be surmised that these uncertainties reflect a real stochastic component in the formation process, which then may be *inferred or bounded* from the data.

This is because a time lag from galaxy formation to actual QSO turn-on of order $t_q \sim$ few Gyr (including hole formation and brightening of the source up to detectability) is important in the range of z envisaged: the parent events would occur at considerably higher z, when all the effective time scales were shorter, satisfying $t(z) < t(2)-t(3)$. If the QSO turn-on constitutes the echo of prior events related to galaxy formation - delayed in time, amplified in luminosity - one may understand why it should be so sharp.

IV. A SPECIFIC MODEL

Let us consider in some detail a simple and specific assumption: bright active nuclei or QSOs originate in protogalaxies of such a high density and total mass as to ensure the collapse of a dense massive core. The two conditions define a range of relevant host masses around some M_g near the leading edge of the galactic luminosity function $N_g(M,z)$, and correspondingly restrict the range of z involved in an average host formation. But the actual distribution of the collapse times corresponding to a given M_g tends to widen the z range for triggering a QSO.

On the other hand, in relating the QSO source term $S(t)$ in Eq 1 to the rate of galaxy formation $S_g(t)$, we have to consider that the QSO turn-on may be delayed from the formation of the host galaxy by a lag t_q; this is in turn subject to a statistical distribution $P(t_q)$ (a sensible representation will be P(t) $\propto exp[-(t-\tau)^2/2\sigma^2]$ with $\tau = \langle t_q \rangle$) to give

$$S(t) = \varepsilon \int_0^t dt' \ S_g(t') \ P(t-t') \qquad \varepsilon < 1 \qquad (2)$$

Once the QSOs are switched on, we assume they undergo a luminosity evolution $\dot{L} = -L/t_f$ that leaves the shape of $N(L)$ unchanged. To characterize the time increase of the QSO population we integrate Eq 1 over L from a lower L_l. Considering that $\dot{L} < 0$ may compete (particularly for the brightest objects) with the change $dN/dt > 0$ we retain the term $N(L_l)L_l/t_f$ set equal to N_q/T_f: this defines a scale $T_f \geq t_f$, the inequality prevailing if L_l lies on the *flat* branch of the luminosity function. Thus we find for the bulk of the QSOs

$$N_q(t) = \int_0^t dt' e^{-\frac{(t-t')}{T_f}} S(t') = \varepsilon \int_0^t dt'' \ S_g(t'') \ \Pi(t-t'')$$

$$with \qquad \Pi(t) = \int_0^t dt' \ e^{-\frac{(t-t')}{T_f}} P(t') \qquad (3)$$

The detailed form of S_g depends on the particular scenario considered, in the present absence of a final understanding of galaxy formation (cf Blumenthal et al 1984). The possibilities range from a self-similar clustering hierarchy (Press and Schechter 1974, Gott and Turner 1977, White and Rees 1978), points of high overdensity (Kaiser 1984), high peaks of density (Bardeen et al 1986) to pancakes (Doroshkevich 1970). But in any case it is hard to avoid the Gaussian nature of the initial fluctuations (Szalay 1985) providing the dominant time scale:

$$S_g(t) \propto \frac{d}{dt}\,[y^{\beta}\ e^{-(\alpha y)^2/2}] \qquad (4)$$

where positive. Here αy describes the typical galactic mass scale collapsing at z: $y \propto t^{-2/3} \propto 1+z$ for $\Omega_0=1$, and more slowly varying in a low density universe (see Cavaliere et al 1977, White and Rees 1978). The quantity α is related to the amplitude of the density fluctuations at some fiducial epoch: For a self-similar hierarchy, $\alpha = (M/M^*)^{(1/2+n/6)}$ with M^* collapsing at a given z, and n the index of the power spectrum for the linear perturbations; in addition, $\beta=1$. For other cases, $\alpha=\nu^2$ in terms of the overdensity ν (normalized to the r.m.s.) then becoming non-linear. With accumulating data and better understanding of QSO formation one may eventually "deconvolve" the source term, providing a strong constraint on galaxy formation.

Figs 2 - 4 focus on the specific model based on Press and Schechter's $S_g = dN_g(M_g,t)/dt$, with $M_g = M^*$ collapsing at $z=2$ or larger. They visualize how simultaneous formation of QSOs and galaxies ($z_g=2$, lag $\tau=0$) would set a scale for $N(z)$ much too slow compared with the fall suggested by the present data. Interestingly short scales are obtained only with $\tau \approx 1$ Gyr for $\Omega_0 = 1$ (most galaxies then would form at $z_g > 5$), or with $\tau \simeq 3$ Gyr for $\Omega_0=0.2$ ($z_g \gtrsim 10$).

V. DISCUSSION AND CONCLUSIONS

The essence of these results is largely model independent. If most of the QSO population is to appear from $z \sim 3$ to 2, dimming or even dying thereafter, the time scale for their increase is considerably shorter than can be the scale $\sim H^{-1}(z)$ for a simultaneous formation of galaxies. A sharp QSO switch-on would be allowed if most proto-hosts collapse considerably *earlier*, at $z_g \gtrsim 5$ (or $\gtrsim 10$ in a subcritical universe). Then the formation era for the parents is confined to a range in cosmic time $\lesssim$1Gyr, in view of two general bounds: the overdense, early collapsing perturbations corresponding to a specific mass are exponentially unlikely; the formation of dense massive cores is least disturbed (White and Rees 1978) by ignition of internal stellar sources at the upper end of the protogalaxy mass range, where $t_{dyn} \sim t_{cool}$ holds (note consistency of a large M_g with the findings of Smith et al 1986). In a hierarchical scenario an additional point helps: the typical mass collapsing at earlier times t decreases like $t^{4/(3+n)}$, faster than $R(t)$.

Very relevant, on the other hand, is the additional dispersion σ corresponding to variance or uncertainties in the process of black hole formation. Figs 2 - 4 visualize that a sharp fall also requires 3σ to be limited to $\lesssim \tau$ for the QSOs. A substantial systematic part of the lag t_q must then shift the actual turn-on from the era of the triggering protogalaxy collapses toward considerably lower z. In other words, the measurement of a real drop in the numbers of QSOs for $z \sim 2$ to 3 will imply the primary hole formation to be a rather long ($t_b \sim 1$

to 3 Gyr, depending on Ω_0) and uniform process; such would be the case for near critical accretion up to $\sim 10^8$ $M_\odot$ with its logarithmic dependence on M.

Larger masses conceivably produce sources that are both stronger and faster to reach detectability, if the final output increase is a runaway process on a time scale depending inversely on L. Once turned on, the ultrabright objects located on the steeper section of $N(L, z > 2)$ easily compensate number increase from $z \sim 4$ to 2 with luminosity decrease, to near invariance of the population with $M_B < -27$.

The QSOs are remarkable for having an evolutionary time scale *shorter* than $H^{-1}(z)$ at all epochs. We have discussed how their apparently sharp turn-on should be linked to substantially *earlier* processes of galaxy formation. Detection of primeval galaxies will restrict the time scales black holes need to form and shine. Otherwise, the emerging QSOs may well be considered a much filtered and distorted echo of galaxy formation, yet with their intrinsic time lag and light amplification they may remain one of the few probes into the parent objects.

It is a pleasure to thank for very helpful discussions R. Ellis, R. Giacconi, D. Koo, A. Renzini and F. Vagnetti. We gratefully acknowledge the hospitality of the Space Telescope Science Institute while this work has been performed. A.S. was supported by the NASA/Fermilab Astrophysics Grant at Fermilab.

FIGURE CAPTIONS

Fig 1. The high-z behaviour of the estimates for: ☆) the number of bright optical QSOs (from Hazard and McMahon 1985); Y) the number of flat-spectrum radiosources (from Peacock 1985). Normalizations to the peak values; high density (flat) universe.

Fig 2. The high-z behaviour of the QSO number $N_q(z)$ as predicted by the model in Sect. IV, for the indicated values of the average lag τ and its variance σ. Points from fig. 1 are replotted for reference. $N_g(z)$ is the normalized Press and Schechter 1974 number of galaxies, with $M_g = M^*$ collapsing at $z = 2$; for $\tau, \sigma = 0$, $N_q(z)$ is just proportional to $N_g(z)$. High density universe.

Fig 3. Same as Fig 2, but for a low density universe.

Fig 4. The fine-grained z-distribution of QSOs corresponding to Fig 2.

REFERENCES

Bardeen,J.M., Bond,J.R., Kaiser,N. and Szalay,A.S. 1986, Ap.J. in press.

Bechtold,J. et al. 1984, Ap.J. **281**,76.

Begelman,M.C. and Rees,M.J. 1978, M.N.R.A.S. **185**, 847.

Blumenthal,G.R., Faber,S.M., Primack,J,R. and Rees,M.J. 1984, Nature **311**, 517.

Cavaliere,A., Danese,L. and De Zotti,G. 1977, Ap.J. 217, 6

Cavaliere,A., Danese,L., De Zotti,G. and Franceschini,A. 1981, Space Sci. Revs., **30**, 101.

Cavaliere,A., Giallongo,E. and Vagnetti,F. 1985, Ap.J. 296, 402.

Condon,J.J. 1984, Ap.J. **287**, 461.

Dahari,O. 1984, A.J. **89**, 966.

De Robertis,M. 1985, A.J. **90**, 998.

De Zotti,G., Boldt,E.A., Cavaliere,A., Danese,L., Franceschini,A., Marshall,F.E. Swank,J.H., and Szymkowiak,A.E. 1982, Ap.J., **253**, 47.

Doroshkevich,A.G. 1970, Astrofizika **6**, 320.

Duncan,M.J. and Shapiro,S.L. 1983, Ap.J. **268**, 565.

Gaskell,M. 1985, Nature, 315, 386.

Gott,J.R. and Turner,E.L. 1977, Ap.J. **216**, 357.

Elvis, M., Wilkes, B.J., and Tananbaum, H. 1985, Ap.J. **292**, 357.

Hazard,C., Morton,C.D., Terlevich, R. and McMahon,R. 1984, Ap.J. **282**, 33.

Hazard,C. and McMahon,R. 1985, Nature, **314**, 238.

Heckman,T.M., Bothun,G.D., Balick, B., and Smith, E.P. 1984, A.J. 89, 958.

Kaiser,N. 1984, Ap.J. **284**, L9.

Keel,W.C., Kennicutt,R.C., Hummel,E. and Van der Hulst,J.M. 1985, A.J. **90**, 708.

Koo,D.C. 1985, in "*The Structure and Evolution of Active Galactic Nuclei*", G.Giuricin, F.Mardirossian eds. (Dordrecht: Reidel).

Koo,D.C., Kron,R.G. and Cudworth,K.M. 1986, submitted to PASP.

McMillan,S.L.W, Lightman,A.P. and Cohn,H. 1981, Ap.J. **251**, 436.

Netzer,H. 1985, Ap.J. **289**, 451.

Norman,C. and Silk,J. 1983, Ap.J. **266**, 502.

Osmer,P.S. 1982, Ap.J. **253**, 28.

Ostriker,J.P. and Heisler,J. 1984, Ap.J. **278**, 1.

Peacock, J.A. 1985, M.N.R.A.S. 217, 601.

Press,W.H. and Schechter,P. 1974, Ap.J. **187**, 425.

Rees,M.J. 1984, Ann.Rev.Astr.Astrophys. **22**, 471.

Roos,N. 1981, Astron.Astrophys. **104**, 218.

Schmidt,M. and Green,R.F. 1983, Ap.J. **269**, 352.

Schneider,D., Schmidt,M. and Gunn,J.E. 1984, B.A.A.S. **16**, 488.

Setti,G. and Zamorani,G. 1984, in "*Advances in High Energy Astrophysics and Cosmology*", COSPAR/IAU Symposium in Rojen.

Shapiro,S.L. and Teukolsky,S.A. 1985, Ap.J. **292**, L41.

Smith,E.P., Heckman,T.M., Bothun,G.D., Romanishin, W. and Balick, B. 1985, preprint.

Szalay,A.S. 1985, *Proc. "Inner Space/Outer Space"*, Fermilab.

White,S.D.M and Rees,M.J. 1978, M.N.R.A.S. **183**, 341.

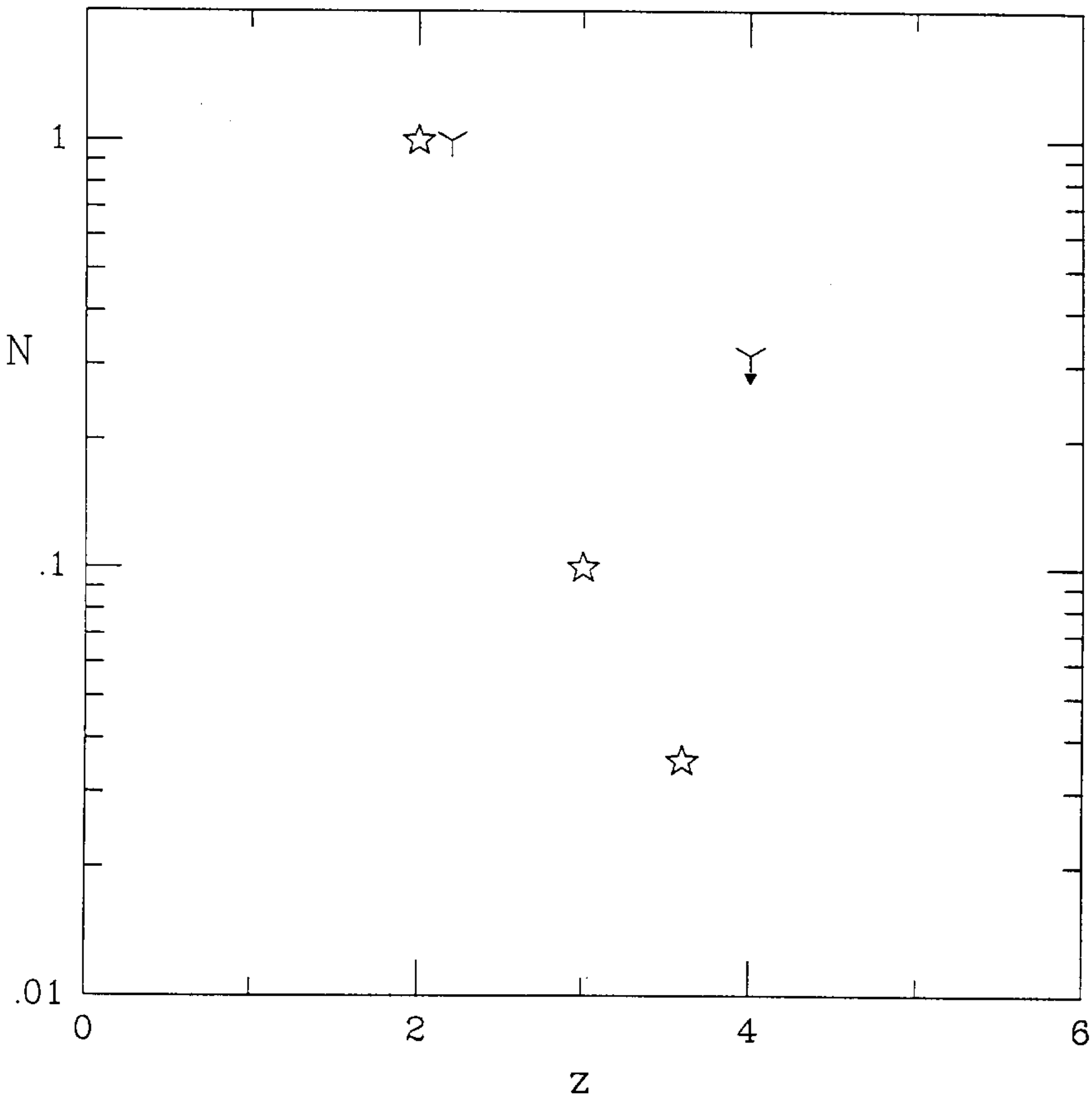
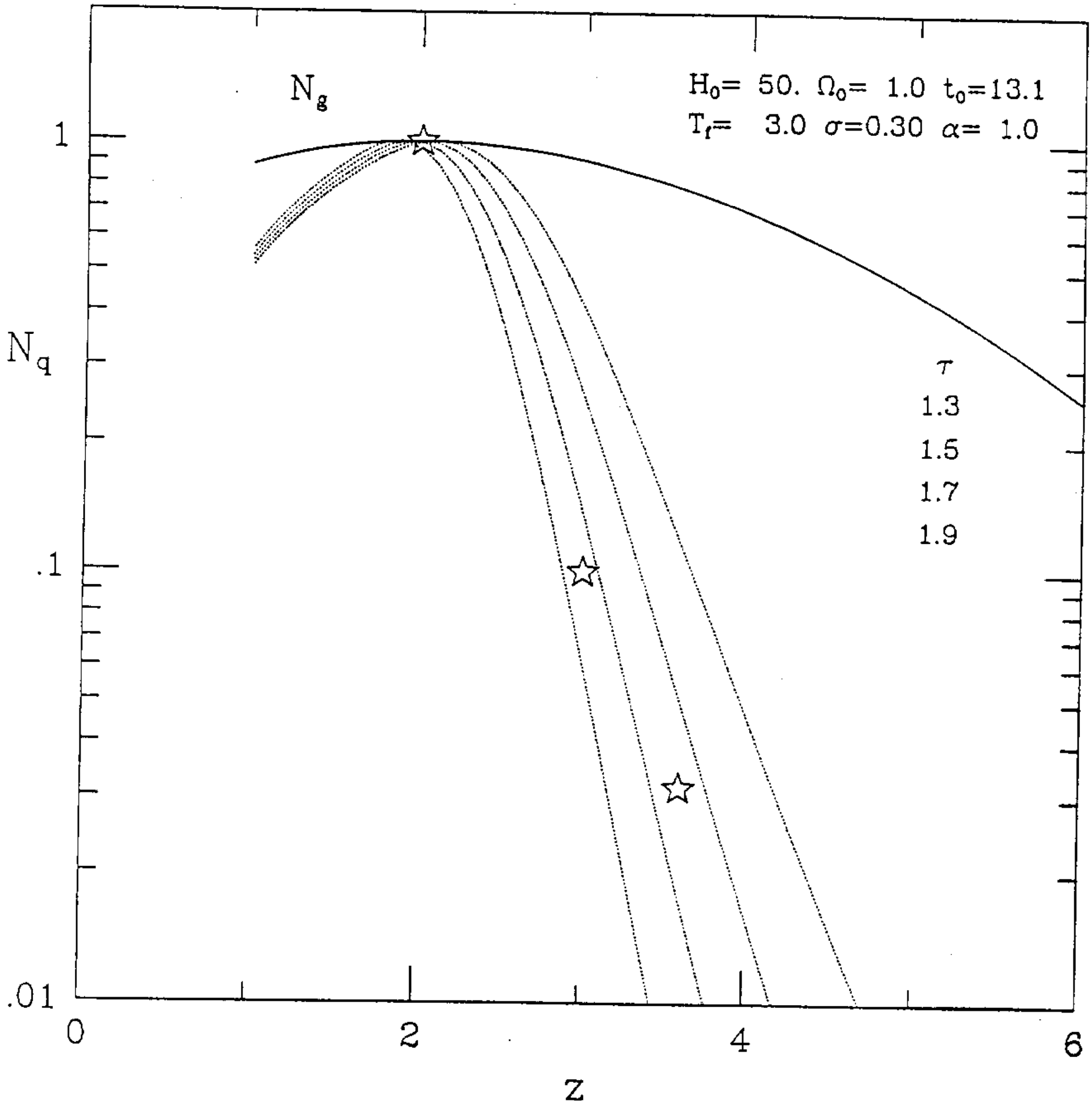


Fig 1



H_0= 50. Ω_0= 1.0 t_0=13.1
T_f= 3.0 σ=0.30 α= 1.0
N_g
N_q
1
.1
.01
τ
1.3
1.5
1.7
1.9
0
2
4
6
z

Fig 2

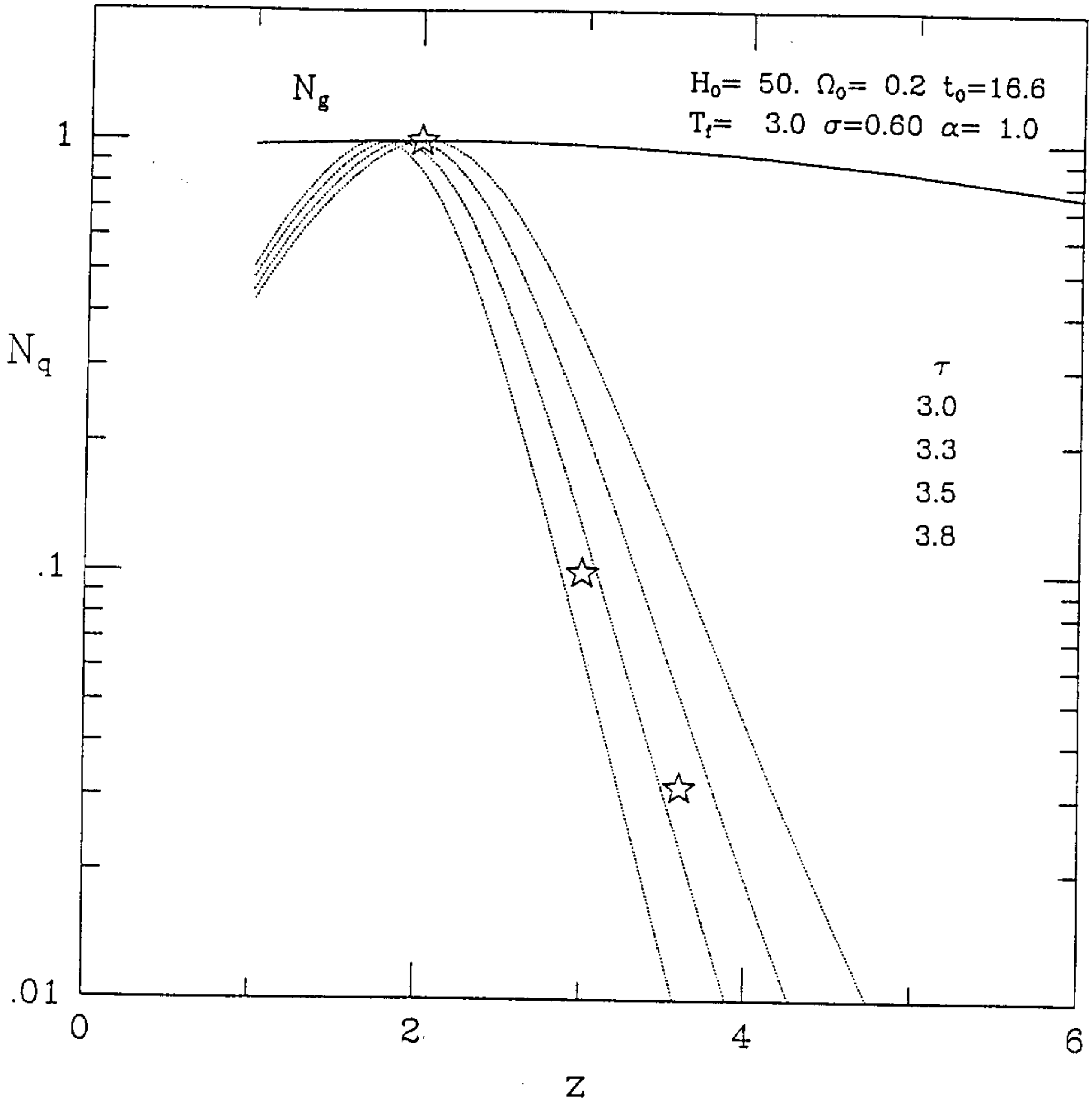


Fig 3

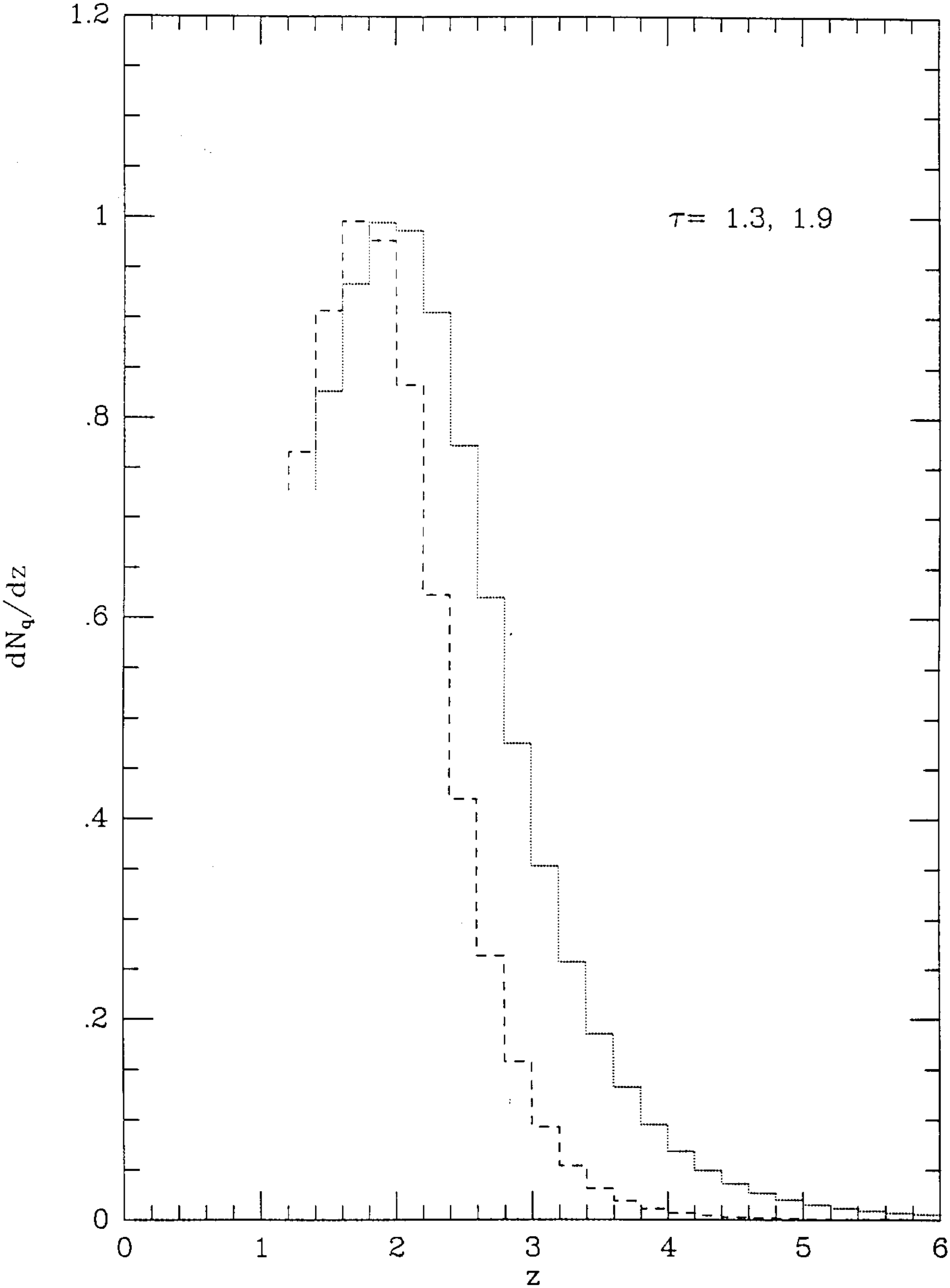


Fig 4

A. CAVALIERE: Astrofisica, Dip. Fisica II Università, via O. Raimondo 8, 00173 Roma, Italy

A.S. SZALAY: Department of Atomic Physics, Eötvös University, Puskin U. 5/7, H-1088 Budapest, Hungary