

The Impact of Galaxy Formation on the Diffuse Background Radiation

Joseph Silk & Julien Devriendt

Oxford University, NAPL, Keble Road, Oxford OX1 3RH, United Kingdom

Abstract.

The far infrared background is a sink for the hidden aspects of galaxy formation. At optical wavelengths, ellipticals and spheroids are old, even at $z \sim 1$. Neither the luminous formation phase nor their early evolution is seen in the visible. We infer that ellipticals and, more generally, most spheroids must have formed in dust-shrouded starbursts. In this article, we show how separate tracking of disk and spheroid star formation enables us to infer that disks dominate near the peak in the cosmic star formation rate at $z \lesssim 2$ and in the diffuse ultraviolet/optical/infrared background, whereas spheroid formation dominates the submillimetre background.

1. Introduction: A Pathway to Galaxy Formation

The theory of galaxy formation became firmly established once temperature anisotropies in the cosmic microwave background were discovered. Only then did we have some indication of the amplitude and power spectrum of the primordial density fluctuations that seeded structure formation via gravitational instability once the universe became matter-dominated. Adiabatic fluctuations in cold dark matter, with the nucleosynthesis baryon component of $\Omega_b h^2 \approx 0.02$ and the cold dark matter component of $\Omega_0 \approx 0.3$, form the basic paradigm for structure formation, supplemented by the recent realisation that $\Omega_0 + \Omega_\Lambda \approx 1$ and $\Omega_\Lambda - \Omega_0 \approx 0.4$ from the CMB $\Delta T/T$ degree-scale and SNIa cosmological probes, respectively.

Inflation suggests that the fluctuations are nearly gaussian distributed, and the problem of galaxy formation, as defined by the assembly of dark halos, is sufficiently well-posed that it has been systematically probed via computer simulations, with varying degrees of success. Highlights of the N-body simulation approach to structure formation include studies of the evolution of the galaxy correlation function, of the mass function and structure of galaxy halos and galaxy clusters, and of the structure of the intergalactic medium. Much of this work has successfully accounted for the observed characteristics of the observed universe. However, not all is in good shape. Galaxy formation to observers is the assembly of the structural components of a galaxy. Predictions of the properties of the luminous components of galaxies have not met with much success. Disk sizes are far too small. The ratio of luminosity to mass is too low. Even the detailed structure of the dark galaxy halos, as probed by high resolution N-

body simulations, is producing results that are discrepant with observations: a density profile with a central cusp $\rho(r) \propto r^{-1.5}$ is predicted that is seen neither in low surface brightness dwarf spirals, barred galaxies, the Milky Way, nor in giant ellipticals. Moreover, excessive substructure is found, leading to apparent conflict with the observed galaxy luminosity function.

Cosmologists have rushed in to resurrect cold dark matter with various alternatives. These range from particle astrophysics solutions that involve, for example, dabbling with the weakly interacting nature of standard CDM, or astrophysical solutions, that invoke recourse to, for example, dynamical feedback of the dissipating baryon matter. While any of these solutions might eventually turn out to have some relevance, a more useful exercise may be to assess the protogalaxy phase of galaxy formation. Improving our understanding of early galaxy evolution is not only closely coupled to observation, but may also help clarify the nature of the protogalactic environment, the role of such astrophysical processes as feedback, and the possible need for more exotic physics.

Ideally, one would like to solve the galaxy formation problem numerically. While with a large N-body simulation, one can identify the sites of galaxy formation, the resolution required to adequately model dynamics alone is at the limit of current computing capabilities. Incorporation of the gas hydrodynamics, required to study star formation, at adequate resolution to cover an entire galaxy, is not feasible. The current limit of simulations only allows study of the formation of the first stellar mass clouds in a protogalaxy, with no possibility of studying either efficiency of star formation or continuing fragmentation. Yet the first generation of stars must have involved millions of massive stars in order to produce the chemical elements seen in the most metal-poor halo stars.

Cosmologists have therefore developed a semi-analytical approach to galaxy formation. This consists of constructing Monte-Carlo realisations of the merging histories of dark matter halos which have collapsed under their own self-gravity. Galaxy formation is then followed by imposing certain simple rules, for example for the rate of star formation. The size of a galaxy is determined by ensuring that angular momentum is conserved in the collapsing cloud. The cloud forms with the small value of initial dimensionless angular momentum $\lambda \approx 0.05$ that is generated by tidal torques between neighbouring fluctuations, and attains virial equilibrium at a fraction λ of the radius at maximum extent. A simple Schmidt law is used for the star formation rate in the resulting disk, and cold gas infall continues to feed the disk. Spiral galaxies are relatively isolated galaxies that have not undergone a merger (that is, of nearly equal mass) within the past 10 Gyr or so. Elliptical galaxies are assumed to form in major mergers on a dynamical time scale: stars are assumed to form in a starburst.

Semi-analytical galaxy formation is able to make realistic maps of the galaxy redshift distribution, and once the feedback parameter is carefully tweaked, the galaxy luminosity function and cosmic star formation history can be derived. However, the model is not very robust, since exploration of the large parameter space for the gas and star formation physics is not really feasible.

It is therefore useful to consider fully analytic models of galaxy formation. We briefly describe such a model in section 2, detail the key procedure which allows us to distinguish between spheroid/disk formation in section 3, and present our results in section 4.

2. Basics of the Model

The cosmological framework we pick here is the canonical Λ CDM model ($\Omega_0 = 1.0$, $\Omega_\Lambda = 0.$, $h = 0.5$, $n = 1$, $\sigma_8 = 0.58$, $\Omega_b h^2 = 0.02$). As discussed by Devriendt & Guiderdoni (2000), only an open universe *without* a cosmological constant would significantly alter our conclusions, and only in the submillimetre window. The results for a flat universe with a positive cosmological constant would not be quantitatively too different from those presented in the following. We then use the Press-Schechter formalism (Press & Schechter 1974), which allows one to compute the comoving number of halos hosting galaxies as a function of redshift, provided the power spectrum of primordial fluctuations in the dark matter density field is known.

To make contact with observations, star formation and chemical evolution must be incorporated. We follow the general framework described by White & Frenk (1991), that is to say, within these halos, we let gas cool radiatively, settle into a disk and form stars. Technical details about the semi-empirical recipes used to account for the different astrophysical processes are given in Devriendt & Guiderdoni (2000). Chemical and spectral evolution models are then computed with the STARDUST model from Devriendt, Guiderdoni, & Sadat (1999). For sake of completeness, we mention that the dust absorption and emission model implemented in STARDUST is the 3-component model of Désert, Boulanger, & Puget (1990), which includes PAHs, very small grains and large grains. This allows us to self-consistently link the optical and the far-IR/submillimetre windows.

We emphasize that, as pointed out in Devriendt & Guiderdoni (2000), there are 3 key parameters in such models: the star formation efficiency β^{-1} , the feedback efficiency ϵ_{SN} and the extent of the gaseous disks f_c . We take parameter values that are fairly similar to the ones used by these authors, with $\beta = 60$, $\epsilon_{SN} = 0.2$ and $f_c = 4$. These values are well within the uncertainties of observations by Kennicutt (1998) and Bosma (1981) for the star formation efficiency and the gaseous extent of cold disks, respectively. Numerical simulations by, e.g., Thornton et al. (1998) tend to give values closer to 0.1 for feedback efficiency; however our higher value is a consequence of an attempt to reduce the number of small objects overestimated by the Press-Schechter prescription, as well as a means to cure the “cooling catastrophe”, where too much gas cools in low mass halos at high redshift. The qualitative effect of each of these parameters can be summed up in the following way:

- Increasing β decreases the normalisation and faint-end slope of the optical and IR counts, because star formation is reduced and takes place at lower redshift.
- Increasing ϵ_{SN} decreases the normalisation and faint-end slope of the optical and IR counts by quenching star formation in higher and higher mass galaxies that form at lower and lower redshifts on average.
- Increasing f_c means increasing the normalisation of the optical counts, and decreasing that of the IR counts, since it is equivalent to reducing extinction.

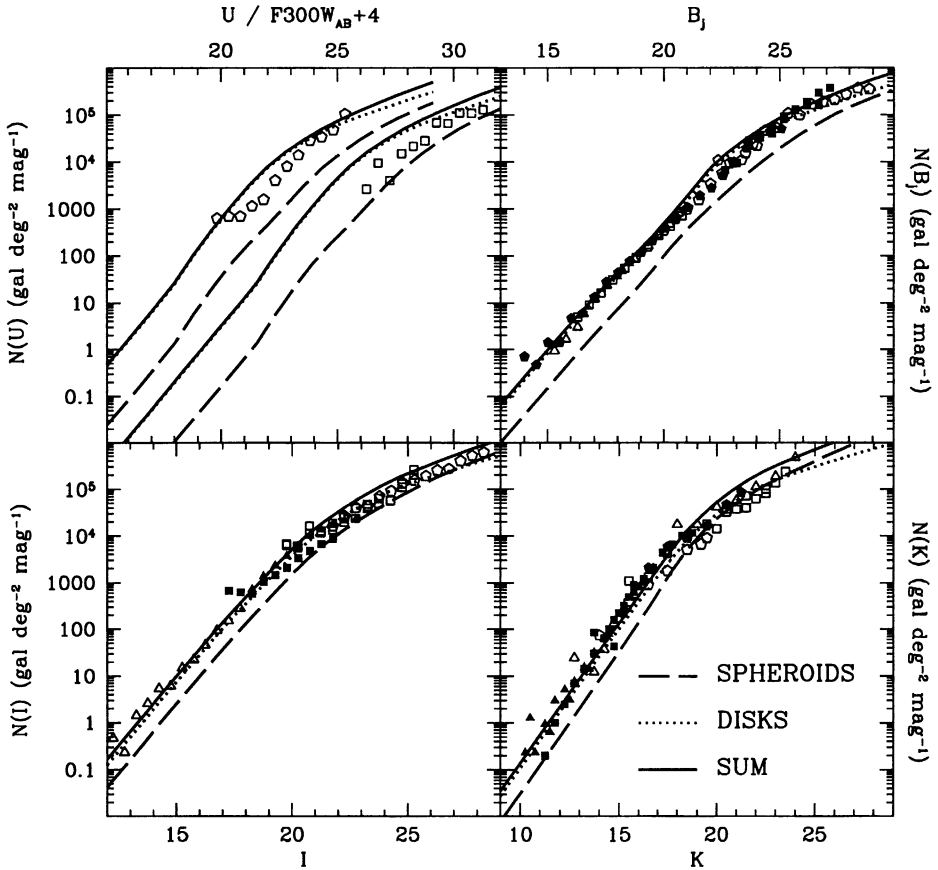


Figure 1. UV/Optical/Near IR counts from our analytic model. Coding for the lines: dotted, late type galaxies (spirals, irregulars); dashed, early type (ellipticals, S0s); and solid, sum of both contributions. Data are from Hogg et al. (1997) (U band), Williams et al. (1996) ($F300W_{AB}$, B & I bands), Arnouts et al. (1997) (B band), Bertin & Dennefeld (1997) (B band), Gardner et al. (1996) (B , I & K bands), Metcalfe et al. (1996) (B band), Weir et al. (1995) (B band), Smail et al. (1995) (I band), Crampton et al. (1995) (I band), Moustakas et al. (1997) (K band), and Djorgovski et al. (1995) (K band).

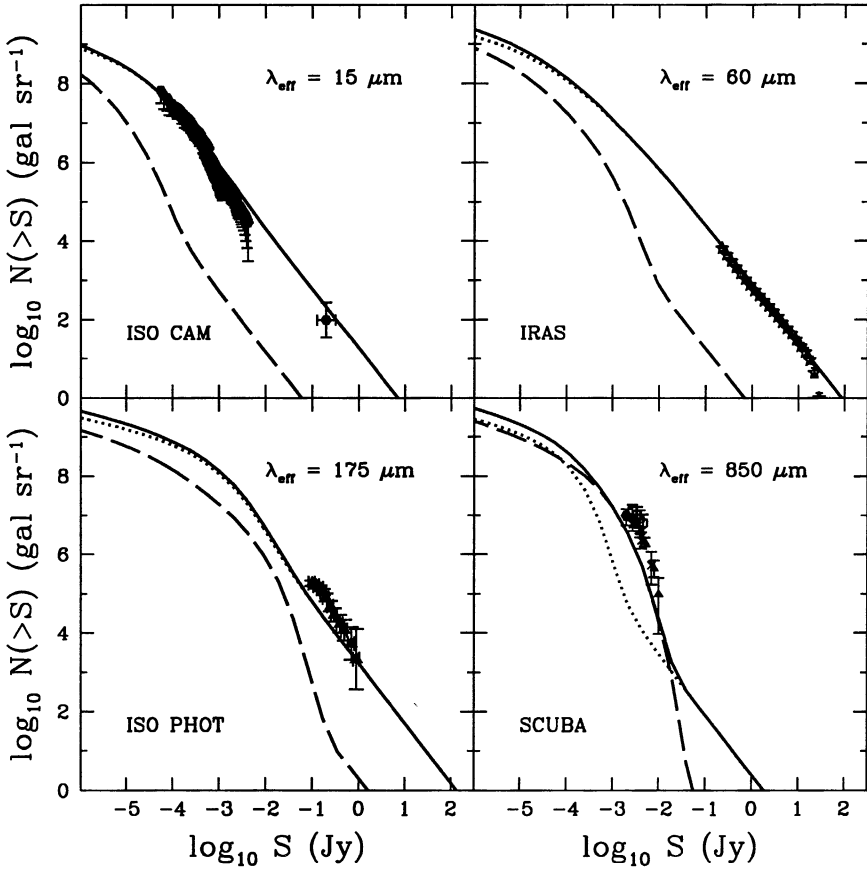


Figure 2. Infrared/submillimetre counts from the analytic model described in the text. Coding for the lines: dotted, late type galaxies (spirals, irregulars); dashed, early type (ellipticals, S0s); and solid, sum of both contributions. Data are from Elbaz et al. (1999) ($15 \mu\text{m}$); Kawara et al. (1998) and Puget et al. (1999) ($175 \mu\text{m}$); Smail et al. (1997), Eales et al. (1999) and Barger et al. (1999) ($850 \mu\text{m}$).

Finally, we point out that the major improvement with respect to Devriendt & Guiderdoni (2000) is the way we model starbursts, by using a physically motivated merger model which allows us to differentiate between spirals and ellipticals. This difference was established in an *ad hoc* way in the model proposed by these authors, where they were simply turning a (redshift-dependent) fraction of the spiral population into ellipticals. Therefore we focus on this issue in the next section.

3. Spheroids Versus Disks

A central issue in galaxy formation is to account for the Hubble sequence of galaxies. The consensus view is that ellipticals form by rare major mergers and that disks form by prolonged infall, which is equivalent to a sequence of minor mergers. In the analytical approach used here, we will use an approximate fit to the collision cross-section to estimate energy exchange ΔE incurred in tidal interactions and mergers. We refer the interested reader to Balland et al. (1998) for a description of how to implement such a calculation within the Press-Schechter formalism.

We then proceed to identify late type galaxies (spirals/irregulars) as undergoing low ΔE , and take early types (ellipticals/S0s) to be defined by high/intermediate ΔE . We define ΔE by integration over galaxy masses and peculiar velocities with number density given by the Press-Schechter formulation. Normalisation of the threshold values of ΔE is achieved by fitting the morphology-density (and also the morphology-cluster radius) relation for rich cluster galaxies. We adopt in the following $\Delta E = 0.003$ as the threshold between late types and early types, which is the threshold used by Balland et al. (1998) for distinguishing between spirals and S0s. The fraction of elliptical, S0 and spiral galaxies can now be predicted as a function of redshift and of mean local density. Field spirals typically form at $z = 1 - 2$, S0s at $z = 2 - 3$, and ellipticals at $z = 3 - 4$.

Depending on the previously identified morphological type, we then decide which galaxies will undergo an “obscure starburst” phase. More specifically, every galaxy identified as an early type galaxy is supposed to go through a LIRG/ULIRG phase, whose intensity and duration are essentially controlled by the amount of gas available for star formation. In practice, this dark phase is modelled by setting our three key parameters to $\beta = 1$, $f_c = 1$ and $\epsilon_{SN} = 0.5$, corresponding to high star formation efficiency, high dust opacity, and high feedback efficiency respectively.

4. Results

The first impression is one of fair overall agreement between model and data, as can be seen from Figures 1 and 2. Looking in more detail, one realises that (except for the faint counts in the near IR bands), late type galaxies dominate over early types (compare the dotted curves with the dashed ones in each panel). This domination extends down to the far-IR, with the late-type galaxies still dominating the $175 \mu\text{m}$ ISOPHOT counts. At longer wavelengths, however, there is a dramatic change: the early-type galaxies completely swamp the contribution from late types. Indeed, one can see on the bottom right panel of

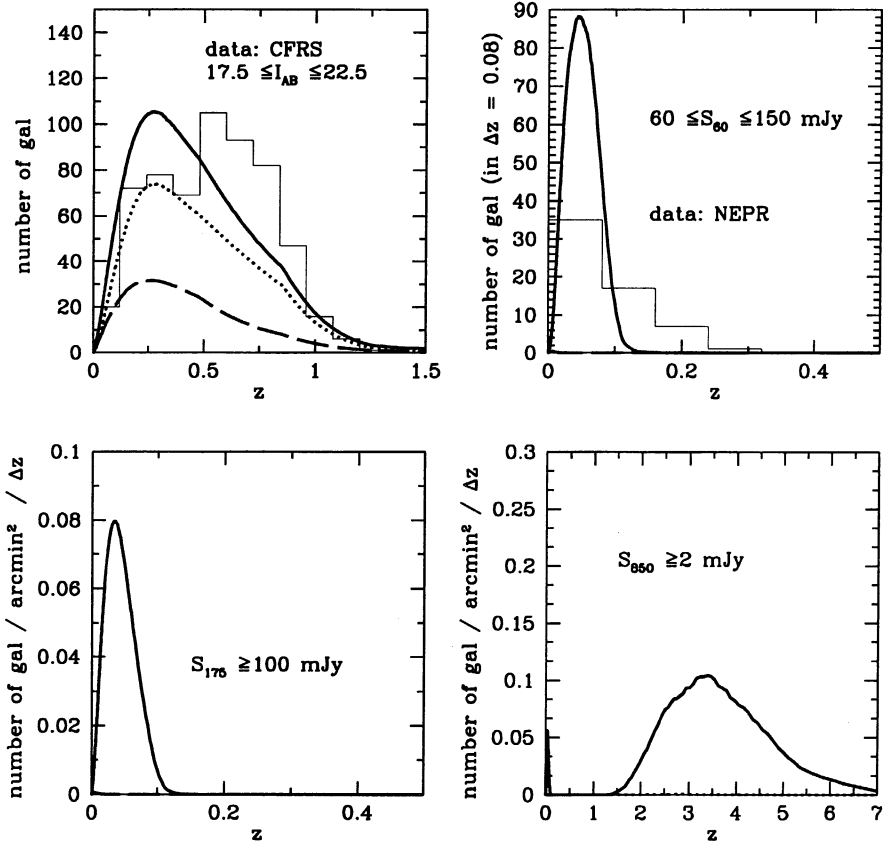


Figure 3. Redshift distributions of the sources in our analytic model. Coding for the lines: dotted, late type galaxies (spirals, irregulars); dashed, early type (ellipticals, S0s); and solid, sum of both contributions. Data are from Crampton et al. (1995) (Canada–France Redshift Survey), and Ashby et al. (1996) (North Ecliptic Pole region). Curves have been renormalised wherever data are shown.

Figure 2 that the vast majority ($\approx 90\%$) of the SCUBA sources are classified as early types in our model. The reason for such a change of behaviour lies in the well known negative k-correction, which makes galaxies of the same bolometric luminosity be as bright for the observer at redshift 5 as at redshift 0.5. This is only important in the submillimetre (here for SCUBA at $850 \mu\text{m}$), because the peak emissivity of dust in these sources' own reference frame is between 60 and $100 \mu\text{m}$. Therefore, as our S0s/ellipticals form at $z > 2$, these maxima of emission must be redshifted to wavelengths greater than 180 and $300 \mu\text{m}$ respectively.

Although this general agreement of our predicted counts with the multi-wavelength data seems quite impressive, there are several caveats. At $15 \mu\text{m}$, for example, one would say that we match the integral counts fairly well (upper left panel of Figure 2). But looking more closely, we cannot reproduce the change of slope seen in the ISOCAM differential counts. There are at least a couple of reasons why this could happen. First, the SEDs of the ISOCAM galaxies might be different in the mid-IR from the ones used here as a template, which are based on IRAS observations of the local universe. This could arise, for example, if the grain size distribution peaks towards smaller grains. Second, because of the way interactions are modelled, each early type galaxy undergoes a starburst after its host halo has just collapsed, and this obviously does not properly describe these sources (see Figure 2), for the vast majority of them are LIRGs (not ULIRGs). Therefore, dynamical interactions (which are not modeled here), triggering multiple milder starbursts, with time delays between them, might provide a more realistic description of these sources. Also, our redshift distributions seem to peak at too low redshifts especially in the *I* band where the mean redshift of the distribution is ≈ 0.4 , whereas the observed peak is more like 0.6 (see top left panel of Figure 3). Moreover, a high redshift tail seems to be missing from the model (top right panel of Figure 3). Both of these difficulties (differential counts at $15 \mu\text{m}$ and redshift distributions) would be eased if a new component of LIRGs is included with a grain distribution whose rest frame emissivity peaks towards $20 \mu\text{m}$. Furthermore, we point out that the redshift distribution is sensitive to the choice of cosmological model: inclusion of Λ would shift the peak towards higher redshifts and also produce a high redshift tail. Finally, we focus the reader's attention on the dramatic change occurring in the redshift distribution of sources when going from ISOPHOT to SCUBA, *i.e.* from the far-IR, to the submillimetre window (bottom panels of Figure 3). Once again, this is due to the negative k-correction boosting observed fluxes in the submillimetre wave-bands.

The background light, as integrated from the multi-wavelength counts, seems to closely match the detection by Puget et al. (1996) and Hauser et al. (1998) in the far-IR/submillimetre window (see Figure 4). This tells us that we are not grossly overestimating or underestimating the faint counts, and that the global galaxy luminosity budget from the UV to the submillimetre is likely to be correctly computed. An interesting remark is that this plot clearly shows that one has to go to the submillimetre, around 750 GHz (about $400 \mu\text{m}$), to see the early-type galaxies dominate over the late types. At any wavelength shorter than this, quiescent galaxies, forming up to $\approx 10 M_{\odot}$ of stars per year, dominate the total light emission (as well as the counts).

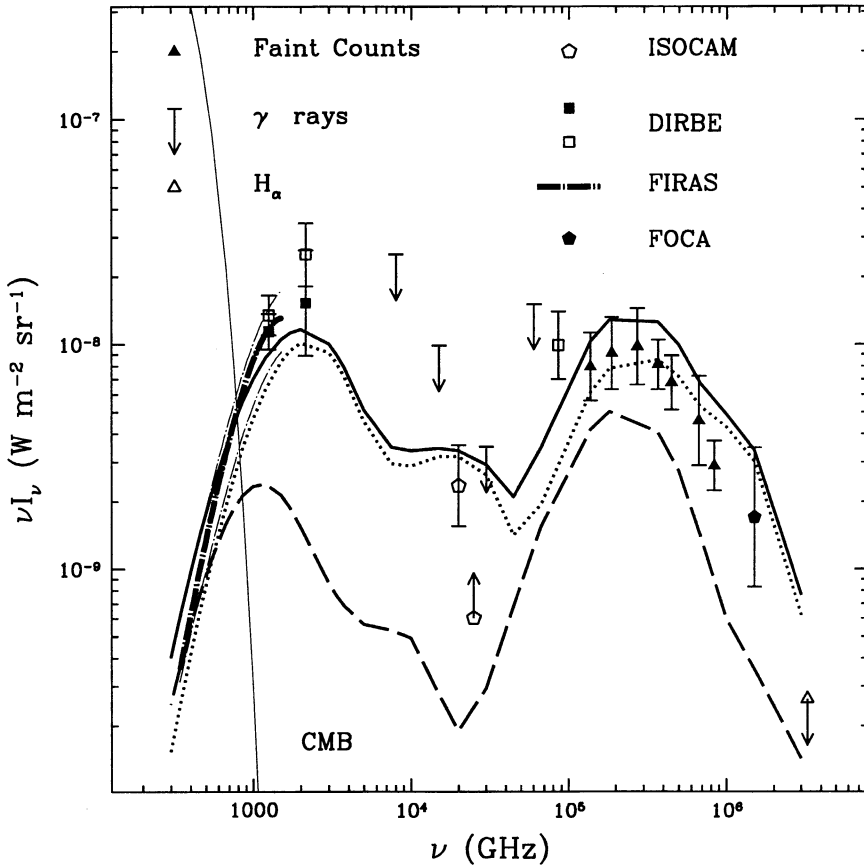


Figure 4. Diffuse background from the UV to the submillimetre as predicted from the model. Coding for the lines: dotted, late type galaxies (spirals, irregulars); dashed, early type (ellipticals, S0s); and solid, sum of both contributions. The data sources are indicated on the Figure.

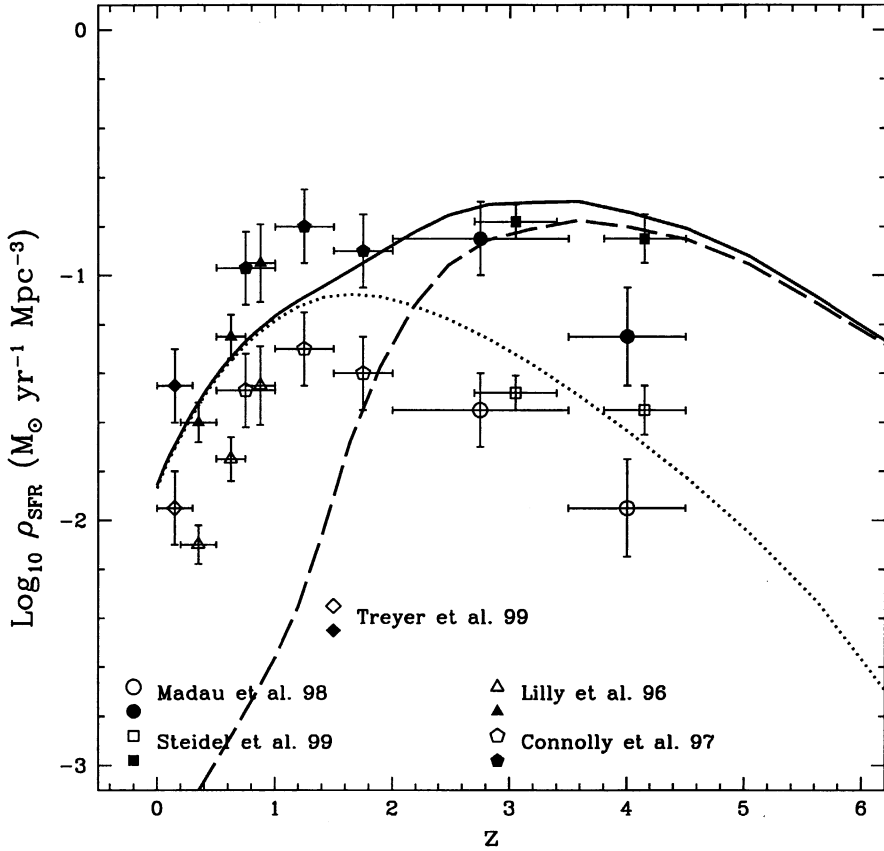


Figure 5. Comoving star formation rate predicted by the model. Coding for the lines: dotted, late type galaxies (spirals, irregulars); dashed, early type (ellipticals, S0s); and solid, sum of both contributions. Empty symbols give raw measurements and filled ones indicate tentative dust corrections (a factor of 3). The data sources are indicated on the Figure.

The comoving star formation rate (Figure 5) is in fair agreement with the measured one, and the reasonable amount of *K*-band light produced (Figure 1) makes us confident that we do not over-produce metals. Furthermore, from this perspective, the star formation rate could be higher, provided that a fair fraction of the metals are ejected into the IGM. The models however seem to point to a peak in star formation around redshift ≈ 3.5 , sensibly higher than what is usually assumed, although current data are compatible with a flat star formation rate at redshifts > 1 . We argue that the amount of star formation occurring in dust-shrouded objects cannot be much greater than our model predicts in order to avoid overestimating the submillimetre diffuse background. Therefore we emphasize that cosmic star formation has to decrease at redshifts ≥ 3.5 . Note that this result is strengthened if dust-shrouded AGNs are a major contributor to the submillimetre emission, even though this does not appear to be the case if X-ray characteristics are a reliable AGN monitor (see Mushotzky, this volume).

5. Conclusions

We have been able to develop an analytical model that enables us to separate the contributions of disk galaxies and spheroids to the panchromatic galaxy counts, the cosmic star formation history and the diffuse background radiation. The quiescent star formation mode characteristic of disks self-consistently accounts for all optical, NIR and FIR data. Only in the submillimetre range does the starburst mode, associated with spheroid formation, dominate. It will be particularly interesting to exploit observations near $400 \mu\text{m}$, where a flux-limited survey should find comparable numbers of forming disks and protospheroids/ellipticals.

References

- Arnouts, S., et al. 1997, *A&AS*, 124, 163
Ashby, M. L. N., et al. 1996, *ApJ*, 456, 428
Balland, C., Silk, J., & Schaeffer, R. 1998, *ApJ*, 497, 541
Barger, A. J., Cowie, L. L., & Sanders, D. B. 1999, *ApJ*, 518, L5
Bertin, E., & Dennefeld, M. 1997, *A&A*, 317, 43
Bosma, A. 1981, *AJ*, 86, 1825
Connolly, A. J., et al. 1997, *ApJ*, 486, L11
Crampton, D., Le Fèvre, O., Lilly, S. J., & Hammer, F. 1995, *ApJ*, 455, 96
Désert, F.-X., Boulanger, F., & Puget, J.-L. 1990, *A&A*, 237, 215
Devriendt, J. E. G., Guiderdoni, B., & Sadat, R. 1999, *A&A*, 350, 381
Devriendt, J. E. G., & Guiderdoni, B. 2000, *A&A*, 363, 851
Djorgovski, S., et al. 1995, *ApJ*, 438, L13
Eales, S., et al. 1999, *ApJ*, 515, 518
Elbaz, D., et al. 1999, *A&A*, 351, L37
Gardner, J. P., Sharples, R. M., Carrasco, B. E., & Frenk, C. S. 1996, *MNRAS*, 282, L1
Hauser, M. G., et al. 1998, *ApJ*, 508, 25

- Hogg, D. W., et al. 1997, MNRAS, 288, 404
 Kawara, K., et al. 1998, A&A, 336, L9
 Kennicutt, R. C. 1998, ApJ, 498, 541
 Lilly, S. J., Le Fèvre, O., Hammer, F., & Crampton, D. 1996, ApJ, 460, L1
 Madau, P., Pozetti, L., & Dickinson, M. 1998, ApJ, 498, 106
 Metcalfe, N., et al. 1996, in IAU Symp. 171, *New Light on Galaxy Evolution*, ed. R. Bender & R. L. Davies, Kluwer Academic Publishers, 225
 Moustakas, L. A., et al. 1997, ApJ, 475, 445
 Press, W. H., & Schechter, P. 1974, ApJ, 187, 85
 Pozetti, L., et al. 1998, MNRAS, 298, 1133
 Puget, J.-L., et al. 1996, A&A, 308, L5
 Puget, J.-L., et al. 1999, A&A, 345, 29
 Smail, I., Hogg, D. W., Yan, L., & Cohen, J. G. 1995, ApJ, 449, L105
 Smail, I., Ivison, R. J., & Blain, A. W. 1997, ApJ, 490, L5
 Steidel, C. C., et al. 1999, ApJ, 519, 1
 Thornton, K., Gaudlitz, M., Janka, H. T., & Steinmetz, M. 1998, ApJ, 500, 95
 Treyer, M. A., et al. 1998, MNRAS, 300, 303
 Weir, N., Djorgovski, S., & Fayyad, U. M. 1995, AJ, 110, 1
 White, S. D. M., & Frenk, C. S. 1991, ApJ, 379, 52
 Williams, R. E., et al. 1996, AJ, 112, 1335

Discussion

Leonid Ozernoy: As you pointed out, the theoretical modeling and a ‘backwards’ approach (to evolve local phenomenology back in time) should come up with consistent results. Meanwhile, a backwards approach by Jim Peebles based on the formation of galaxies by assembling them from smaller ones, leads to the formation of the first galaxies as early as $z \sim 30$, i.e., much earlier than the red shift $z \sim 3$ that you infer for proto-ellipticals. How would you reconcile these results?

Joseph Silk: There is no way that massive galaxies can form as early as $z \sim 30$ for any plausible adiabatic density fluctuation spectrum convolved with gravitational instability theory. Peebles discusses isocurvature seed models, where there is far more freedom for early galaxy formation modelling, but such models are not only theoretically ugly but are virtually excluded by the most recent data on CMB fluctuations.

Charley Lineweaver: How early can you put your proto-elliptical star bursts, $z = 5, 10, 20$?

Silk: Proto-elliptical starbursts can occur at $z = 5$, and perhaps in a Λ -dominated cosmology at $z = 10$. Any earlier than that becomes increasingly implausible given primordial scale-invariant adiabatic density fluctuations, apart from possible early formation of dwarf ellipticals.

Hans Zinnecker: You emphasized the role of star formation in your cosmic evolution picture. I wonder if you can briefly describe how critical the details

of the assumed star formation processes are for your picture, in particular the details of the feedback and the stellar initial mass function.

Silk: Sadly, we have no general theory of star formation. That makes it difficult to extrapolate from very well studied star formation regions, whether Orion or 30 Doradus to such issues as the IMF in the early universe, where physical conditions differed from current conditions. Feedback is a critical assumption in galaxy formation theory. To properly incorporate it, we need a better understanding of whether environment, which might involve outflow from nearby active galaxies, or galaxy collisions, or internal processes, such as supernova-driven winds, dominate. Improved observations of the universe at redshift ≥ 1 , where star formation activity peaks, will help resolve these issues.

Eli Dwek: The formation of the submillimeter background cannot be shifted to very high redshifts; otherwise you would produce an IR background excess at submillimeter wavelengths. Also, the contribution of active galactic nuclei (AGNs) to the cosmic infrared background can be normalized to the observed amount of X-ray absorption as inferred from the turnover in the soft X-ray spectrum of the cosmic X-ray background (XRB). These considerations lead to an AGN contribution of less than 10% at submillimeter wavelengths.

Silk: Our model predicts ULIRG-type starbursts at $z = 3.4$. The theoretical modelling of AGN contributions to the FIRB limits their contributions to at most 30%. This should be regarded as an upper limit that is independent of XRB considerations.

Floyd Stecker: Matt Malkan and I (Malkan, M. A., & Stecker, F. W. 1998, *ApJ*, 496, 13) looked at the observational IR data on galaxies and AGNs and concluded that AGNs make up at most 10% of the IR background. This agrees with what Eli Dwek has just said.

Silk: Using the Magorrian et al. (1998, *AJ*, 115, 2285) relation between black hole mass and spheroid luminosity to predict AGN luminosities required us to assume an efficiency of radiation, as well as a fuelling efficiency. The uncertainties in these normalization parameters mean that we can only limit the far IR AGN contributions to 30% of the IR background.