

## A REDSHIFT SURVEY OF *IRAS* GALAXIES. II. METHODS FOR DETERMINING SELF-CONSISTENT VELOCITY AND DENSITY FIELDS<sup>1</sup>

AMOS YAHIL

Astronomy Program, State University of New York, ESS Building, Stony Brook, NY 11794-2100

MICHAEL A. STRAUSS<sup>2,3</sup> AND MARC DAVIS

Astronomy and Physics Departments, University of California, Berkeley, CA 94720

AND

JOHN P. HUCHRA

Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138

Received 1990 July 16; accepted 1990 October 19

### ABSTRACT

Methods are presented for deriving the density field and the resulting linear-theory velocity field from an all-sky redshift survey of *IRAS* galaxies. The luminosity function of *IRAS* galaxies is measured using a maximum-likelihood technique independent of density inhomogeneities. Within its range of validity, it is found to be in excellent agreement with previously published luminosity functions. Self-consistent solutions for the true spatial positions of the *IRAS* galaxies are found, given their redshifts and peculiar velocities. The ambiguity inherent in triple-valued regions is overcome either by sufficient smoothing, or by using prior information from peculiar velocity maps obtained independently.

*Subject headings:* galaxies: redshifts — infrared: sources — luminosity function

### 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, the peculiar velocity of any galaxy, that is, its velocity relative to the comoving frame, is due to the gravitational pull of the matter around it, and thus should be mirrored in the distribution of this matter.

In linear perturbation theory, initial peculiar velocities are adiabatically damped by the expansion of the universe, and 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 1976)

$$V(\mathbf{r}) = \frac{H_0 f(\Omega)}{4\pi} \int \delta(\mathbf{r}') \frac{(\mathbf{r}' - \mathbf{r})}{|\mathbf{r}' - \mathbf{r}|^3} d^3\mathbf{r}', \quad (1)$$

where  $H_0$  is the Hubble constant,  $\delta(\mathbf{r}') = [\rho(\mathbf{r}') - \rho_0]/\rho_0$  is the mass density fluctuation expressed in units of the average density  $\rho_0$ , and

$$f(\Omega) = d \ln \delta / d \ln a \approx \Omega^{0.6}, \quad (2)$$

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 Equation (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 LG's motion with respect to the comoving frame (e.g., Lubin & Vilella 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.

Equation (1) requires only a dimensionless measure of *relative* density, which can be provided by any set of galaxies which are faithful tracers of mass, at least when averaged on the scales over which density is to be resolved. The combined measurement of peculiar velocity and galaxian density fields thus opens up the exciting possibility of measuring a dynamic value of  $\Omega$ . Moreover, since the number of galaxies involved in the definitions of the velocity and density fields is in each case already in excess of a thousand, 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.

Previous mappings of the local density structure, based on optically selected galaxies (Yahil, Sandage, & Tammann 1980; Davis & Huchra 1982; da Costa et al. 1988; Tully & Fisher 1987) have been constrained in sky coverage and depth. The present series of papers presents the results of a complete redshift survey of galaxies detected by the *Infrared Astronomical Satellite (IRAS)*.<sup>4</sup> We have obtained redshifts for a complete sample of 2649 galaxies, flux-limited at 60  $\mu\text{m}$ , which is selected over 11.06 steradians of the sky. In Paper I of this series

<sup>1</sup> Based in part on data obtained at Lick Observatory, operated by the University of California; the Multiple Mirror Telescope, a joint facility of the Smithsonian Astrophysical Observatory and the University of Arizona; and Cerro Tololo Inter-American Observatory, operated by the Association of Universities for Research in Astronomy, Inc., under contract with the National Science Foundation.

<sup>2</sup> Current postal address: Astronomy Department, California Institute of Technology, Pasadena, CA 91125.

<sup>3</sup> Norris Fellow.

<sup>4</sup> 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).

(Strauss et al. 1990; the present paper is Paper II in this series), we discuss the sample selection and analyze the biases inherent in the *IRAS* database.

This paper describes the methods involved in converting a list of positions and redshifts into a spatial density map, and in calculating the peculiar velocity field generated by this density distribution by linear gravitational instability theory, assuming that *IRAS* galaxies trace the mass. In § 2 the selection function is derived using a maximum-likelihood technique which is independent of density inhomogeneities (Sandage, Tammann, & Yahil 1979, hereafter STY). This is used to correct the observed number density of galaxies for the falloff with distance inherent in a flux-limited sample. The technique is described in §§ 2.1 and 2.2. We justify our basic assumption of a universal luminosity function *a posteriori* in § 2.3 by examination of the luminosity distribution of various subsets of the galaxies. The normalization of the density is discussed in §§ 2.4 and 2.5, and the resulting luminosity function is compared to the work of Saunders et al. (1990) in § 2.6.

In § 3 we use a series of iteration techniques to find self-consistent distances for every galaxy in the sample, such that the observed redshift is given by the Hubble velocity, corrected for the peculiar velocities of the galaxy and the LG as given by equation (1). The assumptions going into this calculation are discussed in § 3.1. In particular, despite evidence that the density structures traced by *IRAS* and optically selected galaxies are not in perfect agreement (Strauss et al. 1991a, Paper IV), we assume that *IRAS* galaxies trace the density distribution of dark matter. In § 3.2 we fill the regions of sky not covered by the redshift survey with random mass points. The calculation of peculiar velocities, and their errors, are described in §§ 3.3 and 3.4. Our methods for finding self-consistent density and velocity fields are outlined in §§ 3.5 and 3.6. A summary follows in § 4. We emphasize that this paper concentrates on *methods* for analyzing the *IRAS* galaxy distribution; the results of this analysis will be presented as follows: Paper III (Davis, Strauss, & Yahil 1991) tests the techniques presented here with *N*-body simulations, and Papers IV, V, and VI in this series (Strauss et al. 1991a, Yahil et al. 1991, and Strauss et al. 1991b, respectively) present the *IRAS* density and velocity fields. Paper VII (Strauss et al. 1991c) will present the data.

## 2. SELECTION FUNCTION

In a flux-limited survey, 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 divide them by the selection function,  $\phi(r)$ : the fraction of the luminosity function included in the sample at a distance  $r$ . The alternative of creating a fully volume-limited sample—including only galaxies luminous enough to be above the flux limit if placed at the outer edge of the volume—rejects so many galaxies from the sample that statistical noise becomes the limiting factor.

A flux-limited sample has only a small number of very sub-luminous galaxies, making it difficult to determine the selection function reliably at small distances. For this reason we impose a lower limit on luminosity,  $L_s$ , semi-volume limiting the sample. Thus, we set the value of the selection function to be unity for  $r < r_s$ , given by  $L_s = 4\pi r_s^2 \nu f_m$ , where  $\nu = 5 \times 10^{12}$  Hz is the frequency of the 60  $\mu$ m band and  $f_m$  is the flux limit.

The results of this section use galaxy distances corrected for peculiar velocities using method 2, described below.  $\Omega = 1$  is

assumed throughout. We have found that the luminosity and selection functions are very insensitive to assumed velocity field, and thus to the value of  $\Omega$  and the method used to calculate the velocity field.

### 2.1. Universal Luminosity Function

A selection function depending only on distance from the observer is meaningless if the luminosity function also depends on environment, particularly if it is a function of density. This is a problem even in a fully volume-limited sample, because there is no guarantee that the fraction of bright galaxies is independent of density. The hypothesis employed here is that of a *universal luminosity function* (Yahil et al. 1980), that is, that 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. This hypothesis needs to be confirmed, at least *a posteriori* (see § 2.3).

Given a universal differential luminosity function,  $\Phi(L)$ , the selection function depends only on distance from us

$$\phi(r) = \frac{\int_{L_m}^{\infty} \Phi(L) dL}{\int_{L_s}^{\infty} \Phi(L) dL}, \quad (3)$$

where  $L_m(r) = \text{Max} \{L_s, 4\pi r^2 \nu f_m\}$  is the minimum luminosity detectable at distance  $r$ . Note that the normalization of the luminosity function cancels out in equation (3).

Our goal, of course, is to convert the spatial number density of galaxies,  $n$ —the raw counts corrected by the selection function—to *mass* density,  $\rho$ . If the infrared galaxies trace the mass, at least when averaged over an adequate volume, then the two densities are proportional. Note that the constant of proportionality is not an issue; we are only interested in *fractional* changes in density, as seen in equation (1). If the *IRAS* galaxies do not trace the mass, that is, if the formation of infrared galaxies is biased, then a prescription is required for transforming one into the other. Linear biasing (Kaiser 1984) is employed here, for which

$$\frac{\delta \rho}{\rho} = \frac{1}{b} \frac{\delta n}{n}. \quad (4)$$

Here  $b$  is the biasing parameter, independent of scale, and equal to unity for the special case when the galaxies trace the mass.

We remark parenthetically that it is possible to measure luminosity density instead of number density. Under the assumption of a universal luminosity function, the two resulting density fields are identical in the mean. In practice, however, the wide range of intrinsic luminosities of *IRAS* galaxies makes the luminosity density a much noisier field.

### 2.2. Density-independent Determination of the Luminosity Function

The old methods to determine the luminosity function of galaxies from redshift surveys (e.g., the review by Felten 1977) are sensitive to spatial inhomogeneities in the galaxy distribution. More modern methods are density independent and use maximum-likelihood techniques; they can be either parametric (STY) or nonparametric (Nicholl & Segal 1983; Efstathiou, Ellis, & Peterson 1988). The method of STY is used here: the probability  $f(L_i | r_i)$  that a galaxy at distance  $r_i$  has the observed luminosity  $L_i = 4\pi r_i^2 \nu f_i$ , is proportional to the value of the

luminosity function at  $L_i$ , divided by the integral of the luminosity function over all possible values of the luminosity it could have at its distance, given the flux limit. Let us define  $\Psi(L)$  as the *cumulative* luminosity function,

$$\Psi(L) = \int_L^\infty \Phi(L') dL'. \quad (5)$$

It is clear from equation (3) that the selection function is

$$\phi(r) = \Psi(L_m)/\Psi(L_s). \quad (6)$$

Thus we can write

$$f(L_i|r_i) = \begin{cases} \Phi(L_i)/\Psi(L_m) & \text{if } L_i \geq L_m \\ 0 & \text{otherwise} \end{cases}. \quad (7)$$

This probability function is density independent; the normalization of the luminosity function drops out in the ratio. Note that  $f$  integrates to unity and is therefore the correct probability function to use in the STY maximum-likelihood fit. For computational convenience we define a likelihood function proportional to the *logarithm* of the product of equation (7) over all galaxies in the sample:

$$\Lambda = -2 \sum_i \ln f(L_i|r_i). \quad (8)$$

The minimum of  $\Lambda$  is arbitrary, but the deviations from the minimum follow the usual  $\chi^2$  rules and provide error estimates for the parameters (e.g., Avni 1976). Given a parameterization for the luminosity function, we minimize  $\Lambda$  for the sample. Strauss (1989) discusses various parameterizations of the luminosity function. Here we use the most robust one

$$\Psi(L) = C \left( \frac{L}{L_*} \right)^{-\alpha} \left( 1 + \frac{L}{L_*} \right)^{-\beta}, \quad (9)$$

with its corresponding *differential* luminosity function

$$\Phi(L) = -\frac{d\Psi(L)}{dL} = \left( \frac{\alpha}{L} + \frac{\beta}{L_* + L} \right) \Psi(L). \quad (10)$$

The resulting selection function is

$$\phi(r) = A r^{-2\alpha} \left( 1 + \frac{r^2}{r_*^2} \right)^{-\beta}, \quad (11)$$

where  $L_* = 4\pi r_*^2 \nu f_m$ , and the constant  $A$  is adjusted so that  $\phi(r_s) = 1$ . Note that our luminosity function is identical to that of Yahil (1988) except in the definition of  $L_*$ , which differs by a factor  $\beta$ .

We minimize equation (8) with respect to the parameters  $\alpha$ ,  $\beta$ , and  $r_*$ , placing galaxies at their redshift distances, corrected to the barycenter of the LG (Yahil, Tammann, & Sandage 1977), and self-consistently for the peculiar velocities predicted from the *IRAS* distribution itself, as described in § 3.5. The semi-volume-limiting radius is set at  $r_s = 500 \text{ km s}^{-1}$ , corre-

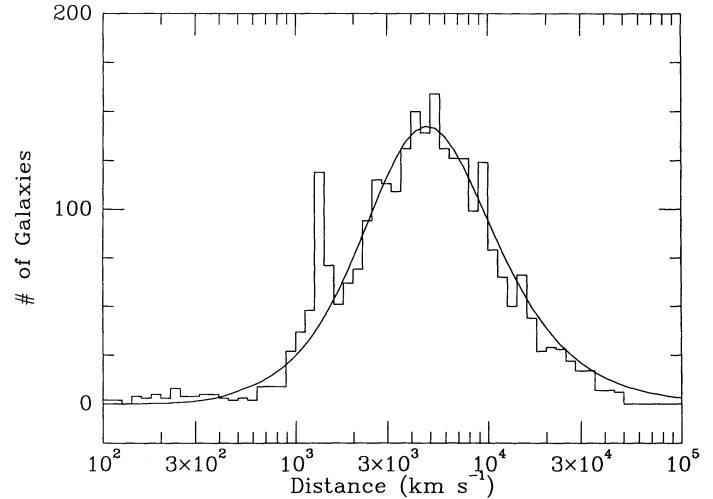


FIG. 1.—Distance histogram of galaxies in the sample, binned in equal intervals of the logarithm of distance. Smooth curve is the expected distribution, given the selection function parameters listed in Table 1. The peak at  $\sim 1200 \text{ km s}^{-1}$  is due to the overdensity associated with the Local Supercluster.

sponding to a low-luminosity cutoff at  $h^2 L_s = 7.4 \times 10^7 L_\odot$ .<sup>5</sup> In addition, because very nearby galaxies may have unknown peculiar velocities which are a significant fraction of their redshift, galaxies with distances less than  $r_s$  are excluded from the fit to the luminosity function, although they are included in the calculation of the density field used in the gravitational calculations below.

At very large distances the selection function is so small, and the weight assigned to each galaxy so large, that statistical noise becomes unacceptable. The determination of density is therefore truncated at a distance  $R_{\max} = 8000 \text{ km s}^{-1}$ . We also exclude these distant galaxies from the fit to the luminosity function. See § 3.1.3 for further discussion and rationalization of the value of  $R_{\max}$ . We find that the resulting selection function is, in any case, quite insensitive to details of the velocity field, the value of  $\Omega$ , and parameters such as  $R_{\max}$ . The parameters determined from the fit, using the full velocity field determined using method 2 below, are listed in Table 1.

Figure 1 shows the distance histogram of the sample, together with the expected distribution in a homogeneous universe, given the best-fit selection function. The small peak at  $\approx 1200 \text{ km s}^{-1}$  is caused by the Local Supercluster. Because *IRAS* covers most of the sky, local inhomogeneities are diluted, and the agreement between the true distance distribution and that expected in a homogeneous universe is very good, especially beyond the Local Supercluster.

There is a slight deficit of observed counts below that expected at distances greater than  $10,000 \text{ km s}^{-1}$ . As we discuss in § 2.5 below, this is because of incompleteness of the

<sup>5</sup> Values of the relative density field,  $(\delta n/n)$ , and the velocity field are both independent of the extragalactic distance scale. To avoid confusion we therefore always quote distances in  $\text{km s}^{-1}$ . 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. As usual,  $h$  is the Hubble constant in units of  $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ .

TABLE 1  
PARAMETERS OF THE LUMINOSITY (SELECTION) FUNCTION

$\alpha$	$\beta$	$h^2 L_*$ ( $L_\odot$ )	$h r_*$ ( $\text{km s}^{-1}$ )	$n_1$ ( $10^{-8} \text{ km}^{-3} \text{ s}^3$ )
$0.527 \pm 0.034$	$1.78 \pm 0.078$	$6.2 \pm 0.65 \times 10^9$	$4400 \pm 230$	$4.3 \pm 0.16$



redshift sample at large distances and low Galactic latitudes (cf. Paper I and Fisher et al. 1991).

### 2.3. Verification of the Universality of the Luminosity Function

The minimization of equation (8) provides confidence limits for the derived parameters but does not test goodness-of-fit. In order to test for that, it is useful to generalize the conditional probability function of a single galaxy to that for a set of galaxies with distances  $r_1, \dots, r_N$

$$f(L|r_1, \dots, r_N) = \sum_{i=1}^N f(L|r_i). \quad (12)$$

As  $f(L|r_i)$  is the probability distribution of luminosity for galaxy  $i$  given its distance  $r_i$ , the quantity in equation (12) is the expected frequency distribution with luminosity of the galaxies in the set. Note that the integral of  $f$  over all luminosities is just the total number of galaxies  $N$ . The observed histogram for a sample may be compared directly with that given by the parameterization of the luminosity function. Any combination of galaxies can be studied using this probability function, as long as the galaxies are not selected by luminosity. It is possible to test for the universality of the luminosity function by considering galaxies at different distances, densities, morphologies, or even colors (provided the colors are available without limits on luminosity), and checking for systematic deviations of the observed and modeled distributions with respect to these parameters.

Figure 2 is a check of the dependence of the derived lumi-

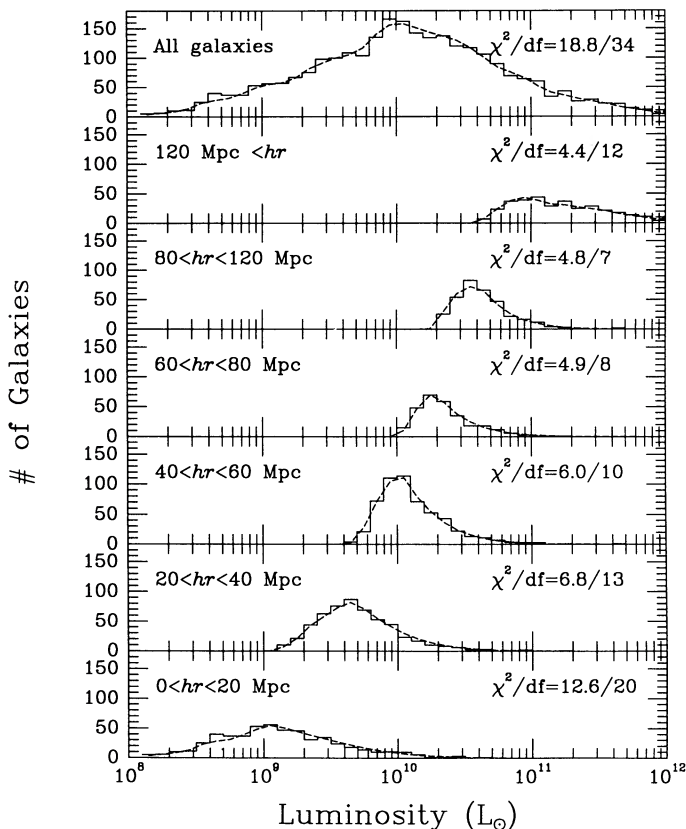


FIG. 2.—Distribution with luminosity of galaxies in different distance ranges, as indicated in the figure. Dashed curves in each case represent the best fit given the parameters of the selection function in Table 1; only the normalization is allowed to vary. Error bars for each point are calculated using Poisson statistics; the resulting values of  $\chi^2$  for each fit are given in the figure.  $\chi^2$  is calculated using only those bins with five or more galaxies.

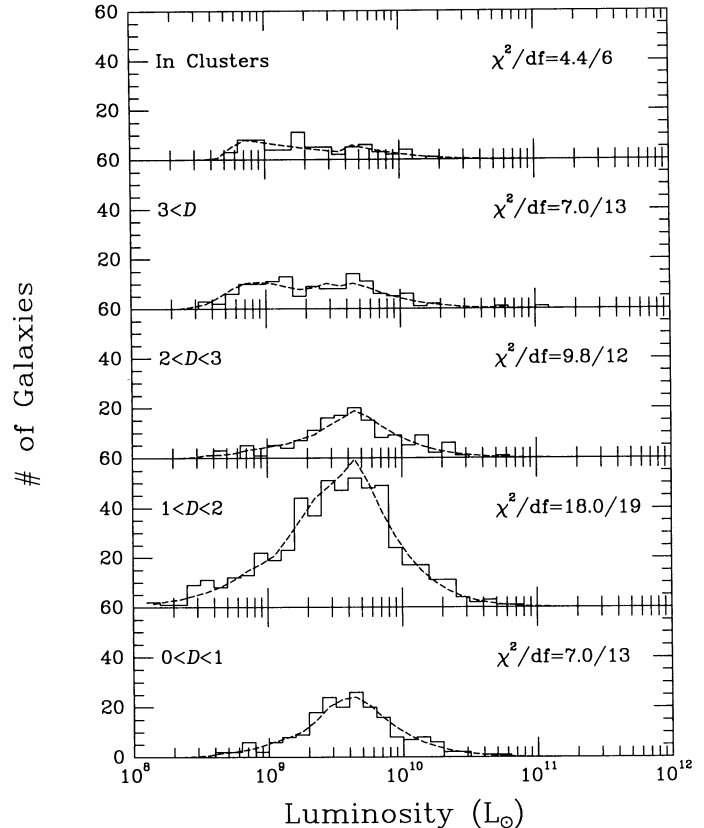


FIG. 3.—As in Fig. 2 for galaxies in different density ranges. Density is calculated using a top-hat smoothing window of  $840 \text{ km s}^{-1}$ , the mean interparticle spacing at the sample limit of  $4000 \text{ km s}^{-1}$ . Top panel shows the luminosity distribution for galaxies in the cores of the clusters in Table 2.

nosity function on distance. The top panel shows the entire sample, while the lower ones break up the counts into distance bins. The dashed lines are plots of equation (12), using the parameters of Table 1. The  $\chi^2$  values have been calculated using error bars given by Poisson statistics and are in the figure; they show that there is no systematic (Malmquist) bias with distance. Moreover, there is an excellent fit to the points beyond  $12,000 \text{ km s}^{-1}$ , over a range of luminosity not covered by the sample to  $R_{\text{max}} = 8000 \text{ km s}^{-1}$ . Our derived luminosity function is consistent with the data to  $60 \mu\text{m}$  luminosities of  $10^{12} L_{\odot}$ . The parameters of the luminosity function are kept fixed to the values listed in Table 1. The only adjustment, implicit in equation (12), is to the normalization.

Similarly, Figure 3 shows a breakup of the observed counts into bins of density relative to the mean. This plot includes galaxies to  $4000 \text{ km s}^{-1}$ , with the density defined by top-hat smoothing by  $840 \text{ km s}^{-1}$ , the mean interparticle spacing at  $4000 \text{ km s}^{-1}$ . Again, the luminosity distribution is fitted with equation (12). The top panel of the figure shows the luminosity distribution of those galaxies assigned to the cores of nearby clusters, as discussed in § 3.5 below. As before, the  $\chi^2$  values show no significant deviations. Finally, we have examined the luminosity distribution of the galaxies broken up into bins of Galactic latitude (not shown here); again, there is no evidence of bias. The luminosity function thus appears consistent with data, providing a *a posteriori* justification for the assumption of a universal luminosity function.

The differential conditional probability, equation (7), can be

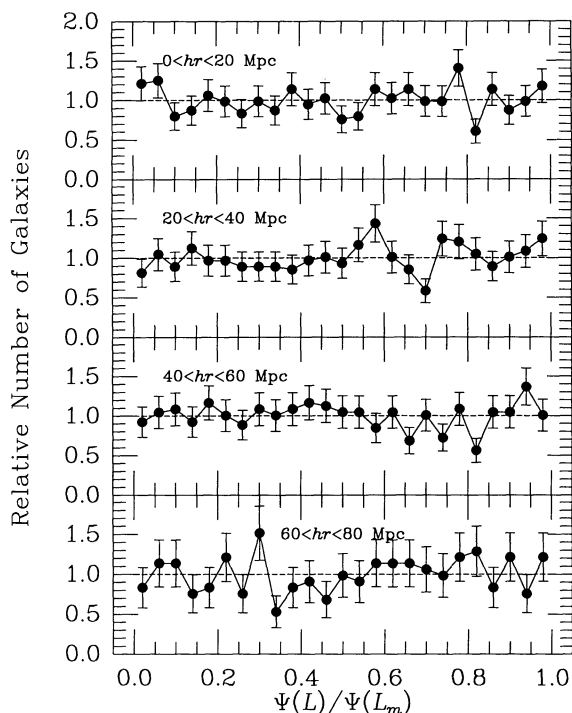


FIG. 4.—Distribution of the quantity  $\Psi(L)/\Psi(L_m)$  in the sample for particles in different distance ranges. The mean value in each case has been normalized to unity. The flatness of this distribution provides *a posteriori* confirmation that the sample is complete.

integrated to yield the cumulative probability  $F(L|r) = \Psi(L)/\Psi(L_m)$ , which should be homogeneously distributed on  $[0, 1]$  independent of distance. The frequency distribution of  $F$  in different distance bins is shown in Figure 4. No deviation is seen from a homogeneous distribution. Note that if the sample were incomplete at the faint end, we would expect a decrease of counts as  $F \rightarrow 1$ . Such a decrease is seen in the Revised Shapley-Ames sample (Sandage & Tammann 1981), leading STY to generalize the formalism used here to a gradual apparent magnitude cutoff function. This is not required for the IRAS sample because of its excellent completeness, as demonstrated by this figure.

#### 2.4. Normalization

In order to measure fractional deviations from uniformity in the galaxy distribution, it is first necessary to estimate the mean number density of galaxies. Several density estimators can be used (Yahil et al. 1980; Davis & Huchra 1982). The one chosen here is

$$n_1 = V^{-1} \sum_i \frac{1}{\phi(r_i)}, \quad (13)$$

where the sum is over all the galaxies in the sample volume  $V$ .  $n_1$  is an estimate of the number density of galaxies with luminosities greater than  $L_s$ . As Davis & Huchra show, this density estimator is unbiased by density inhomogeneities, and it is close to the minimum variance estimator derived in their appendix. For the sample as defined here, namely  $r_s = 500 \text{ km s}^{-1}$  (corresponding to  $h^2 L_s = 7.4 \times 10^7 L_\odot$ ) and  $R_{\text{max}} = 8000 \text{ km s}^{-1}$ , the mean density is  $n_1 = 4.3 \pm 0.16 \times 10^{-8} \text{ km}^{-3} \text{ s}^3$ .

The calculation of density can now be accomplished by assigning each galaxy a weight which is a function of distance

from us,

$$w(r) = \frac{1}{n_1 \phi(r)}. \quad (14)$$

This weight is the mean volume per observed galaxy at distance  $r$ , and the mean observed separation is therefore  $d_{nn} = w(r)^{1/3}$ . In Figure 5 we plot  $w$  and  $d_{nn}$  as functions of  $r$  for the parameters of the selection function listed in Table 1. The mean observed separation varies from  $d_{nn} = 250 \text{ km s}^{-1}$  for  $r < r_s$  to  $d_{nn} = 1800 \text{ km s}^{-1}$  at  $r = 8000 \text{ km s}^{-1}$ .

Given our normalization the density field can be written as

$$D(\mathbf{r}) = \sum_i w(r_i) \delta^3(\mathbf{r} - \mathbf{r}_i). \quad (15)$$

It then follows from equation (13) that the mean density in the entire sample volume,  $V$ , is unity by construction,  $\langle D \rangle = 1$ .

#### 2.5. Sampling Errors in Density

In addition to possible errors in the selection of the IRAS galaxy sample (Paper I), there are two additional sources of error in the determination of density from this sample. First, there is a “shot” noise error due to the finite number of galaxies in any given volume. Assuming that galaxies are uncorrelated samplers of a smooth, but nonuniform, underlying density distribution, the relative error in the estimate of density in any finite volume is, from equation (15),

$$\frac{\Delta D}{\langle D \rangle} = \frac{(\sum_i w_i^2)^{1/2}}{\sum_i w_i} \equiv \frac{1}{N_{\text{eff}}^{1/2}}. \quad (16)$$

This is the generalization of the usual  $N^{-1/2}$  Poisson error to unequal weighting, which can be viewed as due to a smaller “effective” number of sources,  $N_{\text{eff}}$ . Positional correlations among the galaxies on scales smaller than the volume under consideration increase the error even further, since the effective number of galaxies become yet smaller.

The “shot” noise error in density is random, in the sense that it varies from one region to another within the survey volume. A second source of error arises from the determination

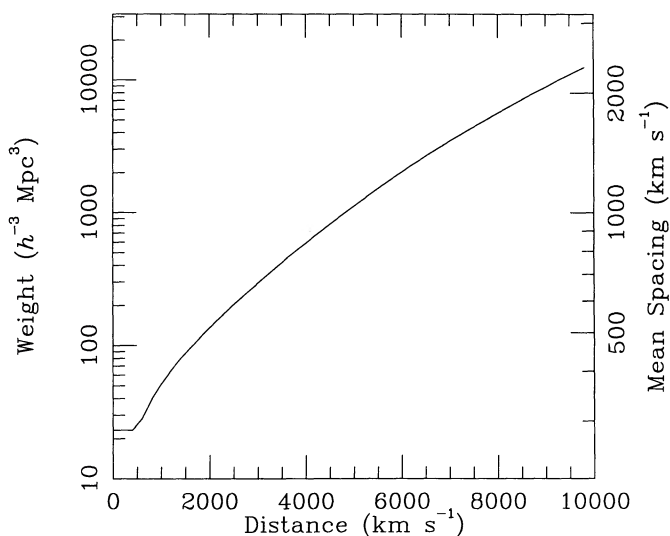


FIG. 5.—Weight function  $[n_1 \phi(r)]^{-1}$  is plotted vs. distance for the sample. Scale on the right-hand side gives the mean interparticle spacing as a function of  $r$ .

of the selection function, however, and introduces a systematic bias with distance via the weighting scheme. In Figure 4 of Paper III we plot  $n_1\phi$ , the inverse weight, as a function of distance for nine random realizations of IRAS catalogs. There is only a small scatter in the weight function at  $r = 4000 \text{ km s}^{-1}$ , about 10%, but there is a much more significant scatter within the semi-volume-limited radius,  $r_s$ , about 50%. More disturbingly, an occasional realization can have an even larger deviation within  $r_s$ . Since  $\phi$  is unity within  $r_s$  by construction, the fluctuation must be due entirely to normalization errors in  $n_1$ .

Davis & Huchra (1982) show that when galaxy correlations are taken into account, the uncertainty expected in  $n_1$  due to inhomogeneities on the scale of the sample is

$$\frac{\Delta n_1}{\langle n_1 \rangle} = \frac{\{\sum_i w_i^2 [1 + n_1 J_3 \phi(r_i)]\}^{1/2}}{\sum_i w_i}, \quad (17)$$

where  $J_3$  is the second moment of the correlation function  $\xi(r)$ ,

$$J_3 = 4\pi \int_0^\infty r^2 \xi(r) dr. \quad (18)$$

As they point out, the minimum variance estimate of  $n_1$  actually occurs for a choice of weight somewhat different from ours. However, the two weighting schemes become identical in the limit  $n_1 J_3 \phi \gg 1$ , which is reasonably well satisfied in our survey volume to  $R_{\max} = 8000 \text{ km s}^{-1}$ . Our weighting scheme, equation (14), is therefore close to optimal in the sense of minimizing the uncertainty in  $n_1$ . We have calculated  $n_1$  using the minimum variance estimate; the result differs from that presented here by less than 1%.

De Lapparent, Geller, & Huchra (1989) point out that if there are appreciable density inhomogeneities on the scale of the sample, the  $J_3$  integral does not converge, and the density estimate is a function of  $R_{\max}$ . We tested this with the IRAS sample as follows. We calculated  $n_1$  from equation (13) using galaxies in logarithmically spaced radial shells from the LG. The selection function parameters found for the  $R_{\max} = 8000 \text{ km s}^{-1}$  sample in Table 1 were used throughout. The result is shown as the solid circles in Figure 6. The error bars are from equation (16). Note that the points in this figure are indepen-

dent. In the range  $3000\text{--}10,000 \text{ km s}^{-1}$ , there are no statistically significant variations in  $n_1$ . However, at larger distances, the estimate of  $n_1$  drops significantly. As Fisher et al. (1991) show, the redshift sample is deficient of high-redshift galaxies at low Galactic latitudes, because it is difficult to identify faint galaxies in regions of high extinction. We tested if this is the cause of the high-redshift drop in  $n_1$  as follows: in each distance range, we calculated  $n_1$  separately for regions above and below  $|b| = 30^\circ$ . The open circles and stars in Figure 6 show the results. The effect of the Local Supercluster is apparent at high latitudes at  $\sim 1000 \text{ km s}^{-1}$ , while the Local Void (Tully & Fisher 1987) depresses the density estimate at lower latitudes. As Paper I shows, the IRAS galaxy sample is complete at high Galactic latitudes, even at high redshift; this is reflected in the stability of the estimate of  $n_1$  from  $3000 \text{ km s}^{-1}$  out to  $30,000 \text{ km s}^{-1}$ . At low Galactic latitudes,  $n_1$  drops precipitously at high redshift. Comparison of the number counts at low and high Galactic latitudes implies that 20–30 galaxies with redshifts above  $10,000 \text{ km s}^{-1}$  have been misclassified as infrared cirrus at low Galactic latitudes (cf. Paper I). These results are consistent with the following statements:

1. There are no appreciable density inhomogeneities averaged over the sky more distant than  $\sim 3000 \text{ km s}^{-1}$  from the observer;
2. The luminosity function models the galaxy distribution well at high redshift (cf. the discussion in § 2.3);
3. The redshift data are not complete at low Galactic latitudes beyond  $\sim 10,000 \text{ km s}^{-1}$ .

## 2.6. Comparison with Other Luminosity Functions

For the purpose of comparison with previous authors, it is useful to consider the number density of galaxies per unit  $\log_{10}$  of  $L$ :

$$\hat{\Phi}(L) = \frac{1}{\log_{10} e} L \Phi(L) = -1.15 n_1 r \left. \frac{\partial \phi(r)}{\partial r} \right|_{r=(L/4\pi\nu f_m)^{1/2}}. \quad (19)$$

The solid line in Figure 7 is  $\hat{\Phi}(L)$ , as given by the best-fit parameters of Table 1 and equation (19). This may be compared with the best-fit parametric luminosity function from

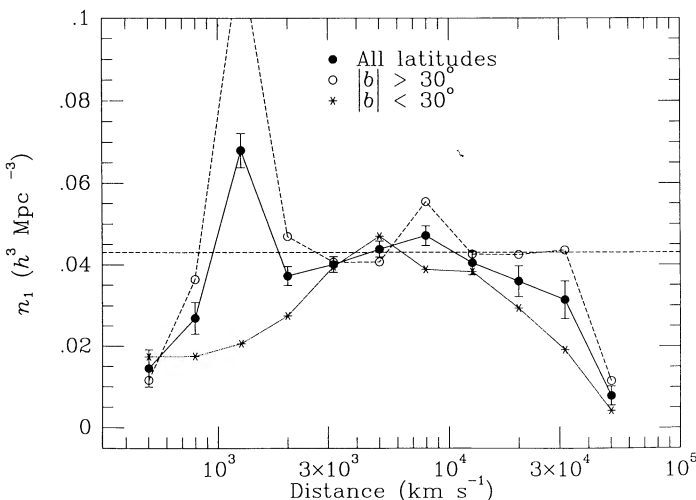


FIG. 6.—Quantity  $n_1$  calculated for independent volumes of space. Solid circles use galaxies at all Galactic latitudes, while open circles and stars are calculated from regions above and below  $|b| = 30^\circ$ , respectively.

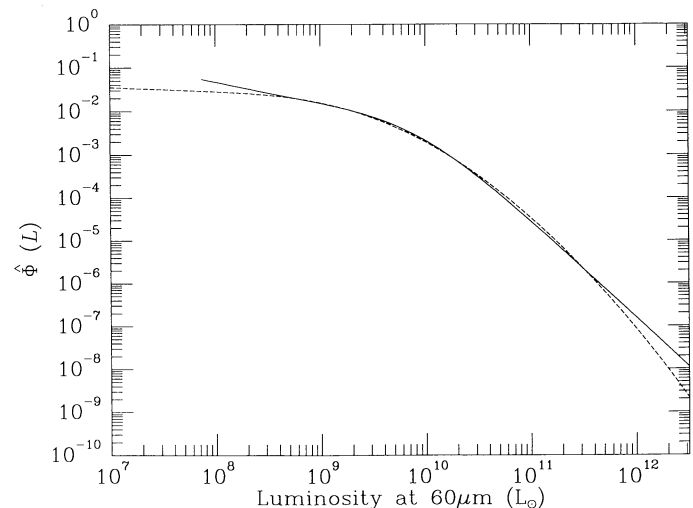


FIG. 7.—Solid curve is the luminosity function given by the parameters in Table 1. The curve is not plotted for luminosities below  $L = L_s$ , where it is not determined. Dashed curve is the parameterized luminosity function derived by Saunders et al. (1990).



seven independent *IRAS* samples by Saunders et al. (1990), which they have shown to be as statistically significant as their nonparametric fit. The two curves agree exceedingly well over their joint range of validity, although the Saunders et al. function is noticeably flatter than that derived here at very low luminosities. The selection function derived here is, by definition, limited to luminosities in excess of  $h^2 L_s = 7.4 \times 10^7 L_\odot$  and is consistent with the data to  $L = 10^{12} L_\odot$ , as we saw in Figure 2. See Strauss (1989) for a comparison with the luminosity function found by the  $V/V_{\max}$  test.

### 3. RECONSTRUCTING THE PECULIAR VELOCITY FIELD IN LINEAR THEORY

In this section we seek to find the peculiar velocity field within the sample volume which is consistent with the measured redshifts and the density field. In this transformation from redshift space to configuration (real) space, we place each galaxy  $i$  with measured redshift in the LG barycenter  $cz_i$  at a distance  $r_i$  such that

$$cz_i = r_i + \hat{r}_i \cdot [V(r_i) - V(0)]. \quad (20)$$

Here  $\hat{r}_i$  is the unit vector in the direction of the galaxy, and  $V(r_i)$  and  $V(0)$  are the values of the peculiar velocity field at the position of the galaxy and the LG, respectively, as calculated by equation (1). We first describe the assumptions that go into this analysis (§ 3.1), then explain how we augment our sample for those regions of the sky not covered by the survey (§ 3.2), and then describe the iteration schemes we use to find the self-consistent solution. These methods have evolved substantially over the last few years and differ in many details from those presented in the conference proceedings of Strauss & Davis (1988a, b; 1991) and Yahil (1988).

#### 3.1. Assumptions

Our determination of the peculiar velocity field relies on a number of assumptions, which should be clearly stated, and subjected to all possible tests.

1. Peculiar velocities of galaxies are caused by gravitational forces, that is, we accept the gravitational instability picture, in which large-scale structure results from initially small perturbations (e.g., Peebles 1980). Other possible sources of peculiar velocities include explosions (Ostriker & Cowie 1981) and cosmic-string wakes (Stebbins et al. 1987). See Peebles (1988) for a detailed analysis of the gravitational instability theory versus strings and explosions as applied to the very quiet velocity field within  $800 \text{ km s}^{-1}$  (Sandage 1986; Brown & Peebles 1987).

The ultimate test is a detailed confrontation of our predicted velocity field with that observed. 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 between them.

2. *IRAS* galaxies trace the matter, as implied by equation (4), on scales larger than the smoothing length used in the calculation of gravity (see § 3.3). They certainly trace the optical galaxy distribution well, as discussed in detail in Paper

IV, except in rich clusters, where spiral galaxies are known to be underrepresented (Dressler 1980). The abundances of spirals, ellipticals, and lenticulars change, however, only at high overdensities  $\delta n/n \gtrsim 100$  (Postman & Geller 1984), and only a small fraction of galaxies are in such high-density regions. All this does not imply that the *IRAS* galaxies trace the mass, because the optical galaxies may not be fair tracers either. This assumption can only be tested dynamically by comparing the density and velocity fields.

3. The gravitational field is dominated by density inhomogeneities inside the volume surveyed by *IRAS*, and inhomogeneities outside it may be ignored. In a multipole expansion of this external gravitational field, the leading term is the dipole, whose strength may be nonnegligible (Vittorio & Juskiewicz 1987; Juskiewicz, Vittorio, & Wyse 1990; Lahav, Kaiser, & Hoffman 1990). This dipole affects the comparison of the *IRAS* prediction of the velocity of the LG with that deduced from the anisotropy of the CBM. The effect of this unknown dipole term can be removed from predictions of the peculiar velocities of galaxies other than our own, 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 a systematic error, 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  $R_{\max}$ . Note also that the transformation from redshift space to configuration space, equation (20), involves only velocities in the LG frame.

The comparison of the *IRAS* predictions with the observed peculiar velocities 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 frames. The advantage of the CMB frame is that the solar velocity relative to it is accurately known from measurements of the CMB dipole, while the solar velocity relative to the barycenter of the LG is uncertain by at least  $50 \text{ km s}^{-1}$  and perhaps  $100 \text{ km s}^{-1}$  (Yahil et al. 1977). When compared with the *IRAS* prediction, however, this disadvantage is offset by the larger uncertainty in the *IRAS* prediction of the LG motion (cf. Paper VI). In this series of papers, we therefore concentrate on the velocity field which *IRAS* predicts in the LG frame.

4. Linear theory holds. *IRAS* galaxies avoid regions of large overdensity, in which linear theory breaks down. In Paper IV we show that *IRAS* galaxies are underrepresented in the cores of rich clusters of galaxies (cf. Babul & Postman 1990). Figure 8 shows a scatter diagram of the relative separation in redshift space of all galaxy pairs, broken into the component along the line of sight,  $\pi$ , and a component on the plane of the sky (projected to a random position angle),  $\sigma$ . In optical galaxy surveys, in which there is a significant contribution of virialized clusters, there is a clearly elongated distribution in the  $\sigma$ - $\pi$  plane ("finger of God") at low  $\sigma$  (e.g., Davis & Peebles 1983), because the peculiar velocities there are greatly in excess of the Hubble separation. No such effect is seen in Figure 8, showing that *IRAS* galaxies tend to avoid high-density regions.

Nonlinear effects can, however, still be important even far from virialized clusters. Several nonlinear corrections have been proposed for spherical perturbations (Yahil 1985; Villumsen & Davis 1986; Regös & Geller 1989; Lightman & Schechter 1990). Yahil (1991) has developed a general nonlinear gravita-

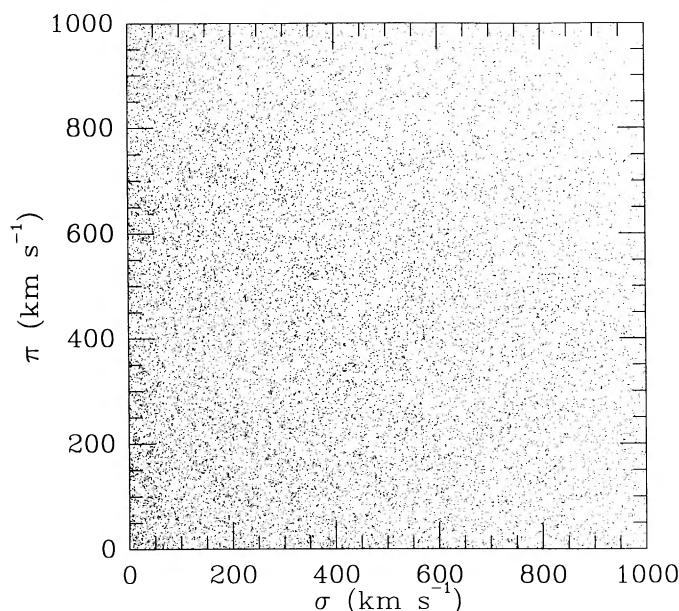


FIG. 8.—Scatter plot showing the breakup of the pairwise separation in redshift space,  $\Delta v_{ij}$ , into a radial (line-of-sight) component,  $\pi$ , and one tangential component,  $\sigma$ . The lack of a “finger of God” concentration of pairs, for which  $\pi \gg \sigma$ , shows that the *IRAS* galaxies avoid virialized clusters.

tional instability scheme, not restricted to spherical perturbations. Its application to the *IRAS* data set seems to be stymied by sampling noise. We are in the process of testing whether this difficulty can be overcome by smoothing, or by the addition of more data in deeper surveys. At this stage we do not use this new nonlinear scheme.

5. The luminosity function of the *IRAS* galaxies is universal. This was tested in § 2.3 and holds very well.

6. There are no residual systematic effects in the *IRAS* sample selection and flux scale. This is discussed in detail in Paper I and is being further tested and improved as we extend the survey to fainter fluxes.

7. The transformation of redshifts to the barycenter of the LG is correctly done. The transformation used here is that of Yahil et al. (1977), but it, like any other transformation, is uncertain by at least  $50 \text{ km s}^{-1}$  and perhaps  $100 \text{ km s}^{-1}$ . This source of error is not expected to be reduced in the foreseeable future.

### 3.2. Unobserved Sky

The *IRAS* redshift survey covers only 87.6% of the sky, and it is necessary to account for the remaining 12.4%. We do this by simulating galaxies in the regions of the sky not included in the survey. These regions fall into two categories: the Galactic strip  $|b| < 5^\circ$ , and high-latitude regions excluded due to insufficient coverage by *IRAS* or by confusion at  $60 \mu\text{m}$  (Paper I). In the latter we sample galaxies from the derived luminosity function, assuming a homogeneous density distribution  $D = 1$ .

In the Galactic strip we wish to have a continuity of structure from one side of the strip to the other, motivated by the fact that both the Perseus-Pisces and Hydra-Centaurus superclusters appear to extend across the plane (Dow et al. 1988; Dressler 1988). We therefore divide the strip into 36 bins of  $10^\circ$  in longitude. In each bin we divide the distance range into bins of  $1000 \text{ km s}^{-1}$ . The galaxies in each longitude/distance bin are then sampled from a homogeneous distribution with density  $D$  equal to the average density in the corresponding longitude/distance bins in the adjacent strips  $5^\circ < |b| < 15^\circ$ . We are indebted to S. Faber for suggesting this procedure; see Lyden-Bell et al. (1989) for a related approach to this problem.

We let the number of artificial galaxies in each volume bin vary as well, being set to a random Poisson deviate whose mean equals the expected number. In this way we allow for statistical fluctuations in both the shape and normalization of the redshift distribution.

Figure 9 shows the sky distribution of all the galaxies in the sample, as well as the random galaxies generated in the excluded zones. Compare this with Figure 18 of Paper I; in particular, notice the coherence of structures across the Galactic plane.

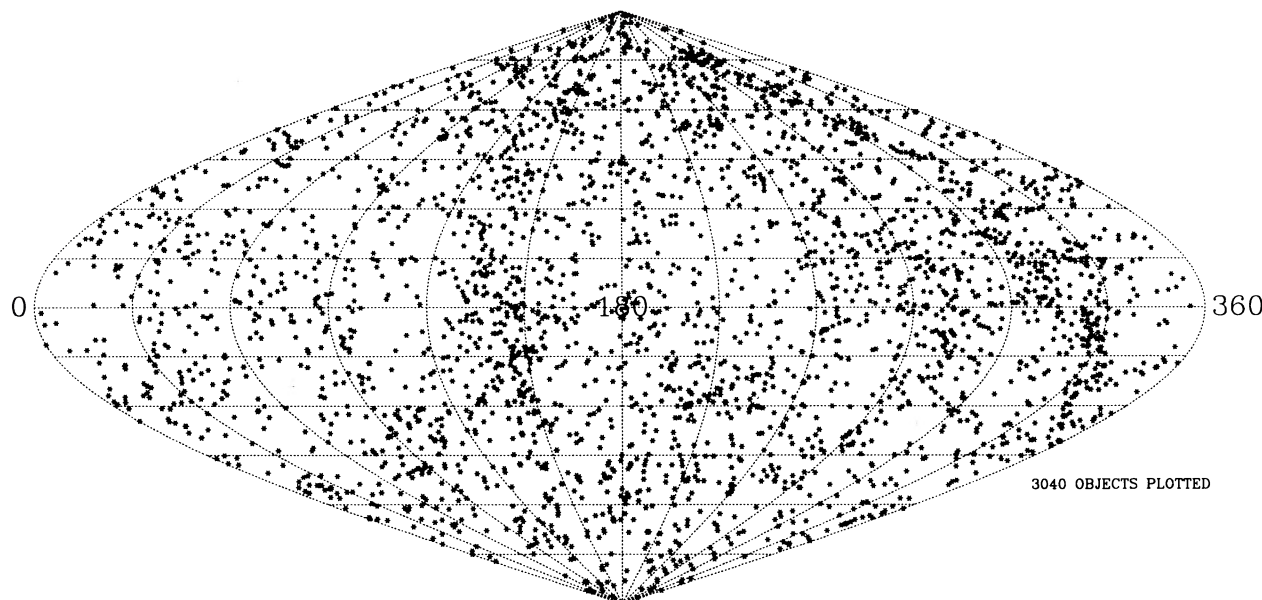


FIG. 9.—Distribution of galaxies in the sky, including those random galaxies generated in the excluded regions. Plot is in Galactic coordinates. Notice continuity of structure across the Galactic plane.



The simulations of galaxies in both the high-latitude regions and the Galactic strip are performed in redshift space after the first fit to the luminosity function and before any corrections are made for peculiar velocities. In subsequent iterations the simulated galaxies are treated as real galaxies: their positions are also corrected for peculiar velocities, and no further simulation is carried out.

### 3.3. Linear Method with Weights

The estimate of the peculiar velocity of a galaxy at position  $\mathbf{r}$  at a distance less than  $R_{\max}$  is given by

$$V(\mathbf{r}) = \frac{H_0 f(\Omega)}{4\pi} \int d^3 r' D(\mathbf{r}') S(\mathbf{r}, \mathbf{r}') \times \frac{(\mathbf{r}' - \mathbf{r})}{|\mathbf{r}' - \mathbf{r}|^3} + \frac{H_0 f(\Omega) \mathbf{r}}{3} + W_S(\mathbf{r}) \hat{\mathbf{r}}. \quad (21)$$

When  $D(\mathbf{r}')$  is given by equation (15), the integral in equation (21) is converted to a sum over all *observed* galaxies within  $R_{\max}$  with luminosities greater than  $L_s$ . Note that  $D(\mathbf{r}')$  is the *total* density of galaxies, not the residual density  $\delta n/n$ . It is therefore necessary to correct for the net gravity of the underlying uniform density field. In the absence of smoothing, the second term on the right-hand side of equation (21) is this correction term. However, as discussed below, a smoothing kernel  $S(\mathbf{r}, \mathbf{r}')$  is also applied, which accounts for the third term in equation (21).

If the galaxy distribution is biased according to the model in equation (4), the quantity  $f(\Omega)$  in equation (21) is replaced with  $f(\Omega)/b$ .

Although only those galaxies within  $R_{\max}$  are taken to be gravitating, we will calculate the peculiar velocities of galaxies at larger distances. For these galaxies, the second term of equation (21) becomes

$$H_0 f(\Omega) \mathbf{r} / 3(R_{\max}/r)^3. \quad (22)$$

The calculation of the gravitational field in the vicinity of another galaxy is problematic for at least four reasons:

1. Equation (21) is derived from linear theory; at small spacings, it no longer holds. This is especially problematic in dense regions, in which galaxies at small spacings give each other large kicks, which are interpreted to mean that their individual distances have to be changed dramatically.
2. Uncertainties in galaxy distances due to peculiar velocities mean that the relative distances between close pairs of galaxies are very uncertain.
3. The gravity calculations are strongly affected by "shot" noise. A single neighboring galaxy can completely dominate the sum in equation (21), especially if  $\mathbf{r}$  is large, in which case the weighting due to the selection function is enormous.
4. A distant galaxy has a large weight and may by itself represent the mass of an entire cluster. It is thus surrounded by an appreciable triple-valued zone, in which the redshift-distance relation becomes nonmonotonic.

With all these problems in mind, we smooth rather heavily on small scales. We need a smoothing kernel  $S$  such that the weight per galaxy is a smooth function of separation  $r$  and falls to zero at small  $r$ . We thus imagine replacing each delta function in equation (15) with a uniform density sphere of radius  $r_{\text{smooth}}$ .

$$S(\mathbf{r}, \mathbf{r}') = \text{Min} \left[ \left( \frac{|\mathbf{r}' - \mathbf{r}|}{r_{\text{smooth}}} \right)^3, 1 \right] \quad (23)$$

We take the smoothing length to be  $r_{\text{smooth}} = 500 \text{ km s}^{-1}$  at small separations. In order to mitigate problem (3) above,  $r_{\text{smooth}}$  is set equal to the maximum of  $500 \text{ km s}^{-1}$  and the average of the mean interparticle spacings at distance  $\mathbf{r}$  and  $\mathbf{r}'$

$$r_{\text{smooth}} = \frac{1}{2} \{ [n_1 \phi(\mathbf{r})]^{-1/3} + [n_1 \phi(\mathbf{r}')]^{-1/3} \}. \quad (24)$$

The smoothing kernel is chosen to be symmetric under the interchange of its variables,  $S(\mathbf{r}', \mathbf{r}) = S(\mathbf{r}, \mathbf{r}')$ , so as to guarantee that the smoothed gravitational forces between pairs of galaxies satisfy Newton's Third Law, and that no artificial bulk flows are introduced by smoothing.

As mentioned above, the second term in equation (21) needs a correction for  $S$ , which is easily calculated to be

$$W_S(\mathbf{r}) = \frac{H_0 f(\Omega)}{4\pi} \int_V dV' [S(\mathbf{r}, \mathbf{r}') - 1] \frac{\mathbf{r}' \cdot (\mathbf{r}' - \mathbf{r})}{|\mathbf{r}' - \mathbf{r}|^3}. \quad (25)$$

The integral extends over the sphere of radius  $R_{\max}$  centered on the origin. The vector  $\mathbf{r}$  in the integral lies in the direction of the  $z$  axis. This function is computed numerically on each iteration. Note that because the smoothing length depends on both  $\mathbf{r}$  and  $\mathbf{r}'$ ,  $W_S$  is nonvanishing even for particles well within the sample volume, for example  $W_S(4000 \text{ km s}^{-1}) \sim 40 \text{ km s}^{-1}$ .

### 3.4. Sampling Errors in the Determination of Peculiar Velocities

As we have emphasized in § 3.1, it is peculiar velocities in the LG frame in which we are most interested, and which we can predict with the greatest precision. Under the same assumptions applied to the density field, namely that galaxies are uncorrelated samplers of a smooth, but nonuniform, underlying density field, the "shot" noise in the radial component of the LG velocity of a source at position  $\mathbf{r}$  due to the finite number of galaxies in the sample, is

$$\Delta[\hat{\mathbf{r}} \cdot (\mathbf{V}(\mathbf{r}) - \mathbf{V}_{\text{LG}})] = \frac{H_0 f(\Omega)}{4\pi} \times \left( \sum_i w_i^2 \left\{ \hat{\mathbf{r}} \cdot \left[ S(\mathbf{r}, \mathbf{r}_i) \frac{(\mathbf{r}_i - \mathbf{r})}{|\mathbf{r}_i - \mathbf{r}|^3} - S(\mathbf{0}, \mathbf{r}_i) \frac{\mathbf{r}_i}{r_i^3} \right] \right\}^2 \right)^{1/2}. \quad (26)$$

Figure 10 plots this quantity for the objects in the sample as a function of distance from the LG. As the test particle nears the sample boundary at  $R_{\max} = 8000 \text{ km s}^{-1}$ , the shot noise is

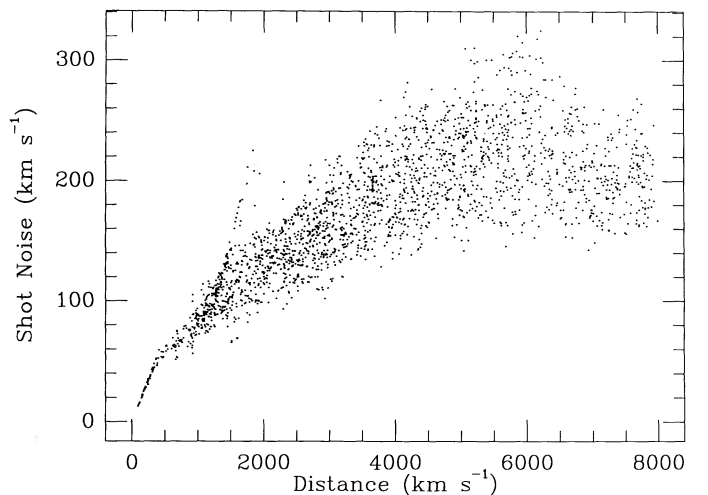


FIG. 10.—Calculated shot noise in the LG-peculiar motion for the galaxies in the sample, as a function of their distance from the origin.

artificially diminished because no noise is associated with the assumption of uniform density beyond  $R_{\text{max}}$ . This causes the turnover at  $\sim 4000 \text{ km s}^{-1}$ .

### 3.5. Iteration Schemes for Finding Peculiar Velocities

#### 3.5.1. Method 1

We are now ready to describe our iteration methods for finding self-consistent solutions to equations (20) and (21). Peculiar velocities are calculated for all galaxies in the sample, including those generated in the excluded regions, but on any given iteration, only those galaxies with distances less than  $8000 \text{ km s}^{-1}$ , and with luminosities greater than  $L_s$ , are taken to be gravitating. Note that we must assume a value for  $\Omega$  in equation (21). In Paper VI we compare the resulting velocity field with the observed field using a variety of values of  $\Omega$ , but for the results of this paper we set  $\Omega = 1$ . To first order, the derived velocity field scales proportionally to  $\Omega^{0.6}$  (cf. eq. [21]), but at smaller values of  $\Omega$  the velocity field does not exhibit the triple-valued behavior that differentiates the methods outlined here.

The following steps are performed prior to the iteration procedure:

Galaxies in the LG (as listed in Binney & Tremaine 1987, Table 10-1) are deleted from the list, so as not to contribute to the gravity on us. There are four galaxies in the LG with  $60 \mu\text{m}$  luminosities greater than  $L_s$ , using the distances listed in Binney & Tremaine and a Hubble constant of  $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ : the Milky Way, M31, M33, and the Large Magellanic Cloud. Thus a point is added to the sample representing the LG, with weight equal to four galaxies. All redshifts are corrected for the solar motion with respect to the barycenter of the LG using the correction of Yahil et al. (1977); the resulting quantity is now called "redshift." Because equation (21) is a linear theory prediction, it is never able to model the "fingers of God" seen in redshift maps of rich clusters. For this reason, we collapse in redshift space the fingers seen in seven prominent low-redshift clusters, listed in Table 2. All these galaxies are assigned to the same angular coordinates as well, as given in columns (1) and (2) of the table. Column (3) gives the adopted mean LG redshifts of the clusters, and columns (4) and (5) give the ranges in redshift and angular radius, respectively, within which galaxies are assumed to be members of the cluster. Finally, column (6) gives the number of galaxies in the sample actually assigned to each cluster. Further discussion of the properties of these clusters in the *IRAS* redshift survey is given in Paper IV.

On the first iteration, each galaxy is assigned a distance equal to its LG redshift; thus the iteration is started with all galaxies comoving with the LG.

Random galaxies are generated in the regions not covered by the survey, as described in § 3.2. After this point they are treated as real galaxies, in the sense that the peculiar velocity on them is calculated on each iteration, and their distances are updated accordingly. These galaxies are *not* included in the selection function determination, however. A total of 393 random galaxies are added. We thus have a total of 3037 galaxies.

At this point we are ready to start the iteration procedure. It is as follows:

1. The selection function is calculated as described above in § 2, using all galaxies with distance  $r_s < r < R_{\text{max}}$ , as calculated on the previous iteration. The density  $n_1$  is similarly calculated.
2. A nested pair of loops is started through all the galaxies; the first loop represents the test particles on which the gravity is calculated, and the inner loop goes through those particles that are gravitating. All the galaxies in the sample are involved in the outer loop, but only those galaxies with distances less than  $R_{\text{max}}$  and with luminosities larger than  $L_s$ , as calculated on the previous iteration, are taken to be gravitating.
3. The peculiar velocity of each galaxy is calculated by equation (21), as is the peculiar velocity of the LG. The acceleration between any two objects within a cluster in Table 2 is set to zero. The radial component of the calculated peculiar velocity is computed. After the first iteration the adopted peculiar velocity for any galaxy is given by the average of the value calculated for the present iteration by equation (21), and that of the previous iteration; this damps oscillatory behavior. Distances are updated using equation (20). Notice that this procedure never makes reference to the observed CMB dipole.
4. The next iteration starts again at step (1) above.

We refer to this procedure as method 1. With differences in detail, it was used in the preliminary *IRAS* reports (Strauss & Davis 1988a, b, 1991; Yahil 1988). Unfortunately, it is inadequate around overdensities, which may be high enough to be surrounded by triple-valued zones, in which the same redshift arises at three distinct distances along a given line of sight. There are two ways to deal with this ambiguity. First, if the density field is smooth enough, that is, if the smoothing kernel  $S(r, r')$  in equation (21) is extended far enough, the density will nowhere rise enough to create triple-valued regions. This may actually be the procedure of choice when comparing with a peculiar velocity field which has been smoothed on a similar scale.

With the current smoothing, there remain appreciable triple-valued zones around clusters, and an arbitration scheme is needed to choose the "best" solution for galaxies in these regions. Method 1, unfortunately, rarely puts a galaxy in the

TABLE 2  
COLLAPSED CLUSTERS IN *IRAS*

$l$ (1)	$b$ (2)	LG Redshift ( $\text{km s}^{-1}$ ) (3)	Length of Finger of God (4)	Radius (5)	Number of Galaxies (6)	Cluster (7)
150°4	−13°4	5559	2000	3°000	7	Perseus
269.6	26.5	3435	1000	3.000	1	Hydra
302.4	21.6	3248	1000	6.000	13	Centaurus
236.6	−54.4	1407	400	3.000	4	Fornax
283.8	74.5	1039	1500	6.000	39	Virgo
144.6	65.5	1011	300	6.000	19	Ursa Major
58.1	88.0	6886	2000	3.000	2	Coma

middle branch of a triple-valued zone, despite the fact that this is the region with largest density, and thus where the galaxy is most likely to be! We therefore replace it with the methods described below.

### 3.5.2. Method 2

To allow the recognition of triple-valued zones, we have replaced step (3) with another step that is somewhat more sophisticated. For each galaxy we tabulate the peculiar velocity for up to forty positions along the line of sight. To avoid increasing the running time of the computations by a factor of 40, we do this in an approximate way. A sphere is drawn around the position of each galaxy, as given on the previous iteration, with a radius equal to 3 times the smoothing length at that distance. The acceleration at this distance due to all galaxies *outside* this sphere is calculated. Then the acceleration due to the galaxies *within* this sphere is calculated at a string of positions along the line of sight separated by  $75 \text{ km s}^{-1}$ . The string is centered on the position of the galaxy and extends  $1500 \text{ km s}^{-1}$  in either direction, unless it reaches the origin or  $R_{\text{max}}$ . The acceleration at each point on the string is then approximately the sum of these inner and outer accelerations. We could then place the galaxy at the distance at which equation (21) predicts the observed redshift. However, this may be ambiguous in the case of triple-valued regions, and it does not recognize the existence of "failed" triple-valued regions, depicted schematically in Figure 11; for a galaxy with the redshift indicated by the dashed line, the solution would in fact become triple valued if the acceleration were modified slightly. The procedure we use instead is more robust: the true distance is taken to be the average of (1) the *minimum* distance along the string at which the predicted redshift is  $200 \text{ km s}^{-1}$  less than the true redshift and (2) the *maximum* distance at which the predicted redshift is  $200 \text{ km s}^{-1}$  more than the true redshift. This process is illustrated in Figure 12, an expansion of Figure 11. The redshift of the galaxy in question is  $1200 \text{ km s}^{-1}$ . Horizontal lines are drawn at  $1000 \text{ km s}^{-1}$  from the left and at  $1400 \text{ km s}^{-1}$  from the right. The positions at which these lines intersect the redshift-distance curve are indicated by

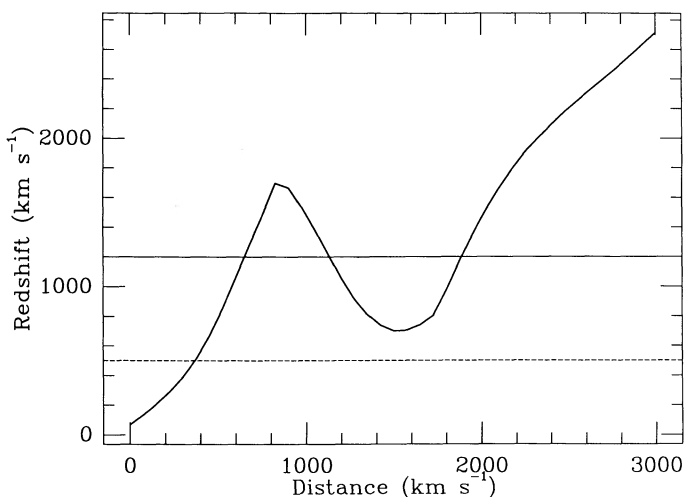


FIG. 11.—Schematic representation of the redshift-distance diagram for a line of sight to a cluster of galaxies. A galaxy with a redshift indicated by the solid line could be at any of three distances, given by the intersection of the line with the curve. A galaxy with a redshift indicated by the dashed line just misses being triple valued.

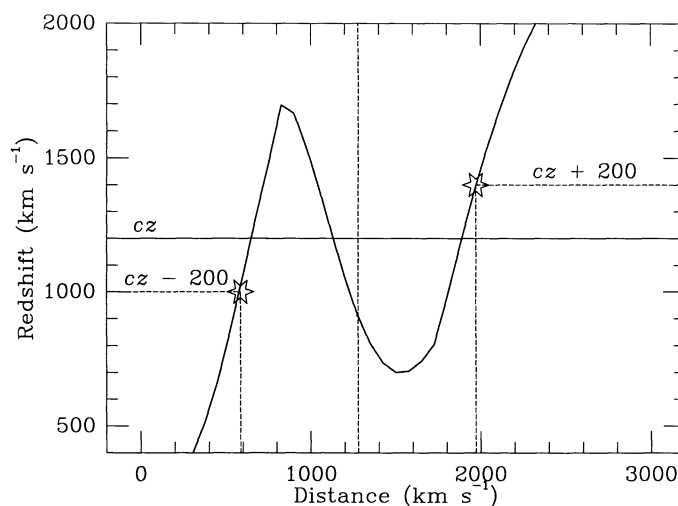


FIG. 12.—Schematic representation of the calculation of distances via method 2. The true redshift is represented by the line labeled "cz." The average of the distances corresponding to the stars is the chosen distance for the galaxy.

large stars. The average of the corresponding distances is then the calculated distance for this galaxy, as indicated by the vertical dashed line.

For galaxies with very small redshifts, there may not exist a distance corresponding to a redshift  $200 \text{ km s}^{-1}$  less than the true value, in which case we revert to method 1, that is, place the galaxy at the distance appropriate to its redshift. The analogous procedure is applied if a galaxy is close to  $R_{\text{max}}$  and the maximum distance is beyond  $R_{\text{max}}$ . The procedure then continues as above: peculiar velocities are averaged between iterations, true distances are computed, and the next iteration starts again at step (1).

As in method 1, the self-gravitation of a point, and the mutual acceleration of points within the same cluster, are suppressed. This need not be done in principle; a test particle along the line of sight to the galaxy in question will feel the gravitational pull of that galaxy. However, we have found that the velocity field coverages much more cleanly, especially in the regions of clusters, when this self-gravity is suppressed.

This procedure is referred to as method 2. Away from the triple-valued regions it gives essentially the same distances as does method 1. In the triple-valued regions, on the other hand, it tends to place all galaxies in the middle branch. This is still not satisfactory, for it does not allow the existence of an infall region around clusters. An unfortunate consequence of this is evident in Figure 13, which shows the predicted positions of all galaxies in the *IRAS* sample within  $22.5^\circ$  of the Supergalactic Plane. The axes are the Supergalactic *X* and *Y* axis, respectively. The point is at the distance of the galaxy, projected onto this plane. The Virgo and Ursa Major clusters are the tight clumps of points at  $(-200, 1000)$  and  $(400, 800)$ , respectively. Because the iteration scheme places all galaxies in the middle solution, it creates a void on either side of these clusters. This is a rather artificial situation; our scheme does not allow for objects in the infall regions of these clusters. Those regions in which this might be a problem are highlighted in Figure 14, in which the galaxies within  $2500 \text{ km s}^{-1}$  are plotted in Galactic coordinates. The size of the symbol for each galaxy is proportional to  $v_{\text{diff}}$ , the difference between the minimum and maximum distances averaged to obtain the true distance for each galaxy on the last iteration, minus  $400 \text{ km s}^{-1}$ . This is a



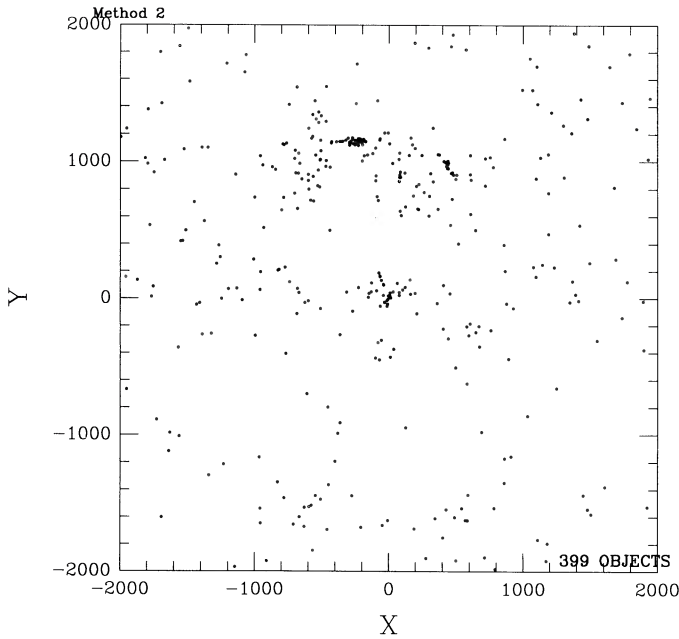


FIG. 13.—Positions of galaxies calculated by method 2 in the Supergalactic plane. All points within  $22.5^\circ$  of the Supergalactic plane are plotted.

measure of the inverse of the slope of the redshift-distance line along the line of sight to each galaxy; the larger  $v_{\text{diff}}$ , the more the velocity field is affected by a strong overdensity. Note the strong delineation of the overdensity associated with the Local Supercluster ridge.

### 3.5.3. Method 3

Unfortunately, peculiar-velocity measurements exist for only a small fraction of the *IRAS* galaxies in our sample, and they suffer from large uncertainties in the distances of individual

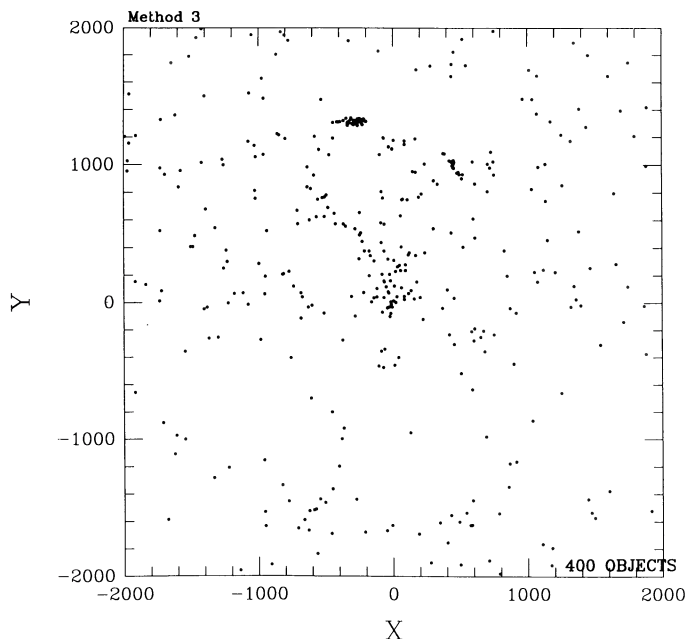


FIG. 14.—Plot in Galactic coordinates of the quantity  $v_{\text{diff}}$  for galaxies within  $3000 \text{ km s}^{-1}$  of the LG. The symbol size is proportional to  $v_{\text{diff}} - 400 \text{ km s}^{-1}$ , which measures the deviation of the redshift-distance relation along the line of sight to that galaxy from pure Hubble flow.

galaxies. We therefore prefer a smoothed version of the observed field to allow us to interpolate to any single galaxy. Dekel, Bertschinger, & Faber (1990) and Bertschinger et al. (1990) have constructed just such a field, weighting the observations by the errors and taking into account the nonuniform sampling of the data and the three-dimensional nature of the true velocity field. The effective smoothing is on a scale of roughly  $1000 \text{ km s}^{-1}$ . There are a number of subtleties involved in this procedure; see their papers for a full description. A. Dekel has kindly supplied us with this field, defined on a rectangular grid with a  $500 \text{ km s}^{-1}$  spacing. Because of the large smoothing length, this field is devoid of triple-valued regions; we can thus unambiguously assign a distance to any galaxy in the *IRAS* sample given its redshift. We can now use this smoothed version of the observed field to iterate the map. We proceed as follows: the velocity field from Bertschinger et al. is used as an initial guess of the true velocity field in the iteration procedure for all galaxies with redshifts less than  $3000 \text{ km s}^{-1}$ . Objects at higher redshifts are assumed to have zero peculiar velocity. This is done as well for the artificial “galaxies” placed in the excluded regions. If, in any iteration, a galaxy is seen to lie in a triple-valued region, the solution which is closest to that given by the smoothed field is chosen. If the solution is single-valued, the program proceeds as in method 2: the true distance is taken to be the average of (1) the minimum distance at which the predicted redshift is  $200 \text{ km s}^{-1}$  less the measured value and (2) the maximum distance at which the predicted redshift is  $200 \text{ km s}^{-1}$  more than the measured value. The difference between methods 2 and 3 is therefore mainly felt in the triple-valued regions. The indirect effect of method 3 outside the triple-valued regions is minimal. The algorithm is not robust enough to recognize the existence of missed triple-valued zones; they are also handled as in method 2. The result is shown in Figure 15 in the same projection as Figure 13; notice that the artificial voids created by method 2 around the Virgo and Ursa Major clusters are filled in by method 3.

Of course, this method is only as good as the measured peculiar velocities that are input into the Dekel et al. (1990) smoothing procedure. If these peculiar velocities have been measured using nonuniversal distance indicators, or are strongly affected by differential Malmquist bias (Djorgovski, de Carvalho, & Han 1988; Silk 1989; Roth 1990; cf. Burstein, Faber, & Dressler 1990), then the results of method 3 are misleading and inaccurate.

As mentioned earlier, if the value of  $\Omega$  is lowered, say, to 0.3 or lower, the velocity field becomes devoid of triple-valued regions, and the results of the three methods become identical.

### 3.6. Convergence of the Iteration Procedure

We have run eight iterations of the velocity field. Convergence of the velocity field is tested by examining the difference in results on the seventh and eighth iterations. There are instabilities in method 2 and 3 associated with the triple-valued regions in the Local Supercluster. We plot the difference of the radial peculiar velocities estimated on the seventh and eighth iteration as a function of distance for all galaxies in methods 2 and 3 in Figure 16. The vast majority of the galaxies have converged to better than  $10 \text{ km s}^{-1}$ , but a few in triple-valued regions bounce from one solution to another in the iterations, in method 3. We are currently examining ways to solve this problem.

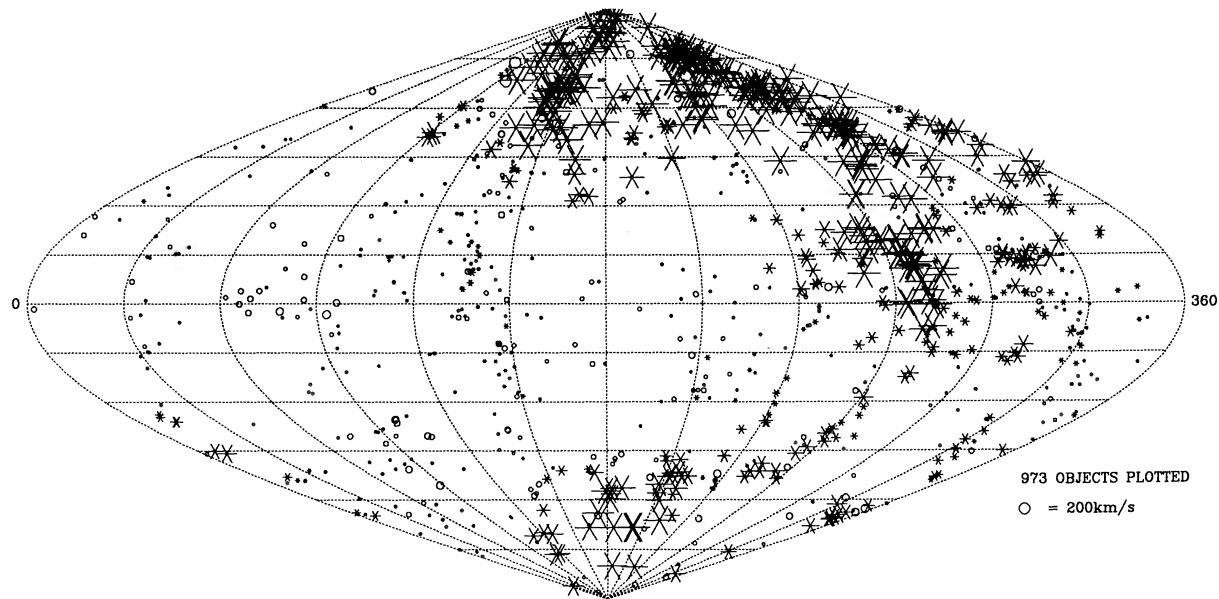


FIG. 15.—As in Fig. 13, for the method 3 results

## 4. SUMMARY

We have discussed methods for obtaining the luminosity function and the linear theory velocity field from an all-sky redshift survey of galaxies. The luminosity function, assumed to be universal, is needed to correct for the decrease in the number density of sources with redshift inherent in a flux-

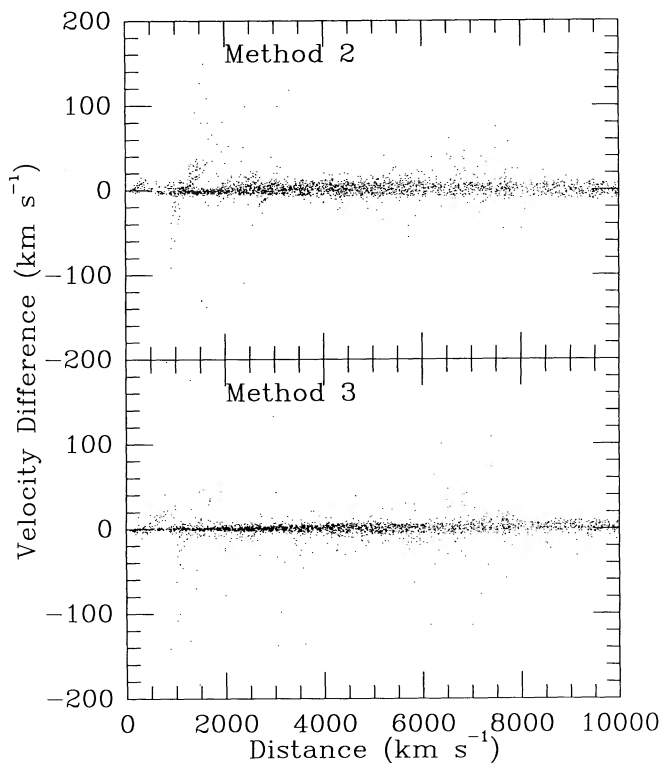


FIG. 16.—Scatterplot of the difference between the radial peculiar velocities derived on the seventh and eighth interactions of methods 2 and 3 for the galaxies in the sample, plotted against their distance.

limited catalog and thus to measure the true galaxy density as a function of position. It is calculated using a maximum-likelihood technique independent of density inhomogeneities in the sample and agrees well with the luminosity function derived by other workers. The normalization of the luminosity function is quite robust to changes in the sample size.

Sparseness of the galaxy sample and nonlinear effects require that we smooth the velocity field on small scales. The presence of triple-valued zones around large mass concentrations means that a self-consistent solution for the velocity field is not unique, at least with  $\Omega = 1$ ; we discuss three methods of increasing sophistication to find that solution most compatible with available observations. In Paper III these methods are tested against Monte Carlo *IRAS* catalogs constructed from *N*-body simulations. Future papers will present the resulting density and velocity fields and will compare them to peculiar velocity observations and results from other redshift surveys.

Depending on the specific application of the methods presented here, the specific details could be changed. An obvious example is the difference between methods 2 and 3. It might be argued that in comparing *IRAS*-predicted peculiar velocities with observations, method 3, with its observational input, biases the results in any comparison between observed and predicted peculiar velocities, especially if the measured peculiar velocities are affected by biases. As another example, we have not justified the specific form of the smoothing function we have used. It gives a discontinuity in slope in the force between two particles, and perhaps a better functional form could be derived. Perhaps also the optimal form of the smoothing may vary, depending on whether one is most interested in calculating the velocity field or the density field. These and other related issues remain to be explored.

We have benefited greatly from discussions on the methods presented in this paper with Ed Bertschinger, Avishai Dekel, Karl Fisher, Andrew Hamilton, Roman Juszkiewicz, Nick Kaiser, Ofer Lahav, Avery Meiksin, Nicola Vittorio, and Rosie Wyse. Josh Roth gave many comments on an earlier draft of this paper that greatly clarified the presentation. This project

has been supported under the *IRAS* extended mission program, grants from the NSF, and NASA grants NAG 51228 and NAG 201. M. A. S. gratefully acknowledges the support of

a National Science Foundation Graduate Fellowship, a Berkeley Graduate Fellowship, and a Norris postdoctoral fellowship.

## REFERENCES

- Aaronson, M., et al. 1989, *ApJ*, 338, 654  
 Avni, Y. 1976, *ApJ*, 210, 642  
 Babul, A., & Postman, M. 1990, *ApJ*, 359, 280  
 Bertschinger, E., Dekel, A., Faber, S. M., Dressler, A., & Burstein, D. 1990, *ApJ*, 364, 370  
 Binney, J., & Tremaine, S. 1987, *Galactic Dynamics* (Princeton: Princeton University Press)  
 Brown, M. E., & Peebles, P. J. E. 1987, *ApJ*, 317, 588  
 Burstein, D., Faber, S. M., & Dressler, A. 1990, *ApJ*, 354, 18  
 da Costa, L. N., et al. 1988, *ApJ*, 327, 544  
 Davis, M., & Huchra, J. 1982, *ApJ*, 254, 437  
 Davis, M., & Peebles, P. J. E. 1983, *ApJ*, 267, 465  
 Davis, M., Strauss, M. A., & Yahil, A. 1991, *ApJ*, 372, 394 (Paper III)  
 Dekel, A., Bertschinger, E., & Faber, S. M. 1990, *ApJ*, 364, 349  
 de Lapparent, V., Geller, M., & Huchra, J. 1989, *ApJ*, 343, 1  
 Djorgovski, S., de Carvalho, R., & Han, M.-S. 1988, in *The Extragalactic Distance Scale*, ed. S. van den Bergh and C. J. Pritchet (ASP Conf. Ser., 4), 329  
 Dow, M. W., Lu, N. Y., Houck, J. R., Salpeter, E. E., & Lewis, B. M. 1988, *ApJ*, 324, L51  
 Dressler, A. 1980, *ApJ*, 236, 351  
 ———. 1988, *ApJ*, 329, 519  
 Efstathiou, G., Ellis, R. S., & Peterson, B. A. 1988, *MNRAS*, 232, 431  
 Felten, J. E. 1977, *AJ*, 82, 861  
 Fisher, K., Strauss, M. A., Davis, M., Yahil, A., & Huchra, J. P. 1991, in preparation  
 Juskiewicz, R., Vittorio, N., & Wyse, R. F. G. 1990, *ApJ*, 349, 408  
 Kaiser, N. 1984, *ApJ*, 284, L49  
 Lahav, O., Kaiser, N., & Hoffman, Y. 1990, *ApJ*, 352, 448  
 Lightman, A., & Schechter, P. 1990, *ApJS*, 364, 831  
 Lubin, P. M., & Vilella, T. 1986, in *Galaxy Distances and Deviations from Universal Expansion*, ed. B. F. Madore & R. B. Tully (Dordrecht: Reidel), 169  
 Lynden-Bell, D., Faber, S. M., Burstein, D., Davies, R. L., Dressler, A., Terlevich, R. J., & Wegner, G. 1988, *ApJ*, 326, 19  
 Nicholl, J. F., & Segal, I. E. 1983, *A&A*, 118, 180  
 Ostriker, J. P., & Cowie, L. L. 1981, *ApJ*, 243, L127  
 Peebles, P. J. E. 1976, *ApJ*, 205, 318  
 ———. 1980, *The Large Scale Structure of the Universe* (Princeton: Princeton University Press)  
 ———. 1988, *ApJ*, 332, 17  
 Postman, M., & Geller, M. 1984, *ApJ*, 281, 95  
 Regös, E., & Geller, M. J. 1989, *AJ*, 98, 755  
 Roth, J. 1990, preprint  
 Sandage, A. 1986, *ApJ*, 307, 1  
 Sandage, A. J., & Tammann, G. A. 1981, *Revised Shapley-Ames Catalogue of Galaxies* (Washington: Carnegie Institute of Washington)  
 Sandage, A., Tammann, G. A., & Yahil, A. 1979, *ApJ*, 232, 352 (STY)  
 Saunders, W., Rowan-Robinson, M., Lawrence, A., Efstathiou, G., Kaiser, N., Ellis, R. S., & Frenk, C. S. 1990, *MNRAS*, 242, 318  
 Silk, J. 1989, *ApJ*, 345, L1  
 Stebbins, A., Veeraraghavan, S., Brandenberger, R., Silk, J., & Turok, N. 1987, *ApJ*, 322, 1  
 Strauss, M. A. 1989, Ph.D. thesis, University of California, Berkeley  
 Strauss, M. A., & Davis, M. 1988a, in *IAU Symposium 130, Large Scale Structures of the Universe*, ed. J. Audouze, M.-C. Pelletan, & A. Szalay (Dordrecht: Reidel), 191  
 ———. 1988b, in *Large Scale Motions in the Universe: A Vatican Study Week*, ed. V. C. Rubin & G. V. Coyne S.J. (Princeton: Princeton University Press), 256  
 ———. 1991, in *Large Scale Structures and Peculiar Motions in the Universe*, ed. D. Latham & N. L. da Costa (ASP Conf. Ser.), in press  
 Strauss, M. A., Davis, M., Yahil, A., & Huchra, J. P. 1990, *ApJ*, 361, 49 (Paper I)  
 ———. 1991a, *ApJ*, submitted (Paper IV)  
 ———. 1991b, in preparation (Paper VI)  
 Strauss, M. A., Huchra, J. P., Davis, M., Yahil, A., & Tonry, J. 1991c, in preparation (Paper VII)  
 Tully, R. B., & Fisher, J. R. 1987, *Nearby Galaxies Atlas* (Cambridge: Cambridge University Press)  
 Villumsen, J. V., & Davis, M. 1986, *ApJ*, 308, 499  
 Vittorio, N., & Juskiewicz, R. 1987, in *Nearly Normal Galaxies*, ed. S. M. Faber (New York: Springer), 451  
 Yahil, A. 1985, in *The Virgo Cluster of Galaxies*, ed. O. G. Richter & B. Binggeli (Garching: European Southern Observatory)  
 ———. 1988, *Large Scale Motions in the Universe: A Vatican Study Week*, ed. V. C. Rubin & G. V. Coyne, S.J. (Princeton: Princeton University Press), 219  
 ———. 1990, in preparation  
 Yahil, A., Sandage, A., & Tammann, G. A. 1980, *ApJ*, 242, 448  
 Yahil, A., Tammann, G. A., & Sandage, A. 1977, *ApJ*, 217, 903  
 Yahil, A., et al. 1991, in preparation (Paper V)