THE ASTRONOMICAL JOURNAL VOLUME 84, NUMBER 10 OCTOBER 1979

A SURVEY OF GALAXY REDSHIFTS. I. DATA REDUCTION TECHNIQUES[a)]

JOHN TONRY and MARC DAVIS[b)]
Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138
Received 15 May 1979; revised 21 June 1979

ABSTRACT

We are currently undertaking a magnitude limited redshift survey of galaxies having $m_B \leq 14.5$, $b^{II} \geq 40$, $\delta \geq 0$. In this paper, we present in some detail our methods of data reduction, which are based on cross correlation against filtered templates. We present expressions for the uncertainty of a measured redshift, for the internal broadening of the object, and for the uncertainty of this broadening. Comparison of our optical data with previously published 21 cm data shows no systematic errors and yields excellent agreement with our internal error analysis. The method of analyzing velocity dispersions is new and quite promising for further application. A series of spectra are presented as examples to show the power and limitations of the correlation techniques.

I. INTRODUCTION

In the past several years there has been much interest and activity regarding redshift surveys of galaxies. The Shapley–Ames catalog of the entire sky to $m_B \simeq 13.0$ is now nearly complete (see Sandage 1978); many observers have chosen sections of the sky to probe deeper. Chincarini and Rood (1975), Tifft and Gregory (1976), Tarenghi *et al.* (1979), and others have measured large numbers of redshifts in the directions of several rich clusters. Kirshner, Oemler, and Schechter (1978) have obtained nearly complete redshift data to $m_B \simeq 15.5$ in eight small fields in the north and south galactic polar caps. Huchra *et al.* (1979) have nearly completed the survey of the Gott–Turner sample galaxies ($m_B \leq 14.0$, $b^{II} > 40$, $\delta > 0$), which contains 1084 objects. A group of us at the CFA are completing this survey and extending it to $m_B \leq 14.5$, a total of 1942 galaxies.

The motivation for these surveys is diverse. Low-quality data (i.e., velocity uncertainty $\geq$ 100 km/sec) is adequate for determination of the luminosity function and space distribution of galaxies, but high-quality data (i.e., velocity uncertainty $<$ 50 km/sec) are desirable for dynamical studies. Most of the data in the literature have large uncertainty and are therefore of limited utility for accurate measurements of Ω, the cosmological density parameter. Recent 21 cm data have in general yielded velocity measurements with very small errors.

Our motivation for undertaking a redshift survey is similarly multifold. Because the angular galaxy covariance function w (θ) for the 14.5 m sample scales properly to deeper samples, Davis and Geller (1976) conclude it may begin to be a "fair sample" of the universe. If this is true, then Ω determined for such a sample may be a reasonable approximation to Ω for the homogeneous universe. Such a sample of galaxies would be appropriate for intercomparison with the N body studies of Aarseth. The sample would be invaluable for studies of cosmography on large scales and could result in better understanding of the initial conditions leading to the present galaxy distribution.

Statistical methods of determining Ω from redshift surveys have been discussed by numerous authors (e.g., Geller and Peebles 1973, Fall 1975, Davis, Geller, and Huchra 1978, and Peebles, 1979). These methods will work reliably if the root-mean-square (rms) measurement error of velocity is considerably less than the rms peculiar velocity of galaxies. The latter could range from 50–1000 km/sec, depending upon whether a particular galaxy is associated with a poor or a rich cluster.

The bias of any determination of Ω will be minimized if the correction for measurement error is itself minimized. Therefore, we are attempting to determine accurate redshifts (rms error <50 km/sec) for the galaxies in our sample, and equally important, we seek a realistic estimate of the expected errors of our redshifts. Because some of the already published data have large uncertainty, or poorly determined uncertainty, we shall in due course reobserve many of the brighter galaxies in the Northern Sky.

Early type galaxies are not easily detectable in 21 cm radiation and there are occasional misidentifications because of source confusion in radio telescopes. Careful analysis of optical spectra can yield very accurate redshifts, and this paper details what we feel to be a near optimal algorithm for spectral analysis, which is a fully linear process easily programmable on a minicomputer and suitable for processing a large quantity of data.

[a] Research supported in part by NSF Grant AST76-22675
[b] Alfred P. Sloan Fellow

 0004-6256/79/101511-15$00.90

Cross-correlation techniques have been applied by numerous authors beginning with Simkin (1974); a recent thorough review has been given by Davies (1978). Brault and White (1971) have detailed the proper methods of Fourier analysis of spectral data, which we shall use extensively throughout this discussion.

This paper will discuss our procedures for collection and analysis of galaxy spectra. The instrumentation and operation at the telesclope is described briefly in Sec. II. Section III describes in some detail the theory of cross-correlation analysis, including error estimation. We have realized that the width of the cross-correlation peak is a good measure of the intrinsic spectral broadening of a source and in Sec. III we explore the power and uncertainties of this new method of dispersion measurement. We believe it compares favorably with Schechter's method (Sargent *et al.* 1977, hereafter SSBS) for determination of velocity dispersions, but our data base for comparison with known dispersions is presently small. Although our primary intent remains the compilation of a redshift catalogue, many of our data are of sufficiently high quality to yield a useful measure of the intrinsic broadening, and therefore this information will be assembled into our final catalogue.

Section IV describes the stages of the actual data reduction procedure with a number of figures showing data for M81 at various stages of processing. Section V gives a pictorial presentation of the correlation analysis of typical data, some with excellent signal-to-noise, and some with very poor signal-to-noise. Section VI presents comparison studies of the residuals of our measurements against previously published high-quality 21 cm redshifts and against published optical redshifts. The agreement is excellent, indicating that we do understand the major sources of error in our data.

II. OBSERVATIONAL TECHNIQUE

We obtain medium dispersion spectra using the CFA Photon Counting Reticon Detector on the Mt. Hopkins 60″ telescope. The detector is a 936 by 2 diode Reticon fiber optically coupled to a high-gain image-tube package. The electronic processor is copied from the design of Shectman (Shectman and Hiltner 1976). Details of the operating characteristics of this device have been given by Davis and Latham (1979). Our typical resolution is 30–40 μ full-width half maximum (FWHM) and we generally use a 600 line/mm grating in first order, giving a dispersion of 120 Å/mm, or a resolution of 4.8 Å. Photon events are located in position to $\frac{1}{2}$ diode (one pixel) and so we obtain simultaneously two 1872 pixel spectra directly in digital counts which are recorded on floppy disk. The diodes of the Reticon are 375 by 30 μ with no dead space; with the demagnification of the spectrograph, the projected separation between the two spectra is 30″ on the sky. We have been using a pair of slits 12.5″ by 3.2″ for most of our observations. With this slit width, the resulting spectral resolution is only slightly degraded from the detector resolution.

The detector has poor response to the blue of 4000 Å because of a fiber optic input window on the first photocathode, so our usual grating tilt covers a span of 2600 Å, from 4300 to 6900 Å. With the grating in this position, we can measure Hα and N II emission from sources with $v < 15{,}000$ km/sec, and at the same time we measure the strong absorption lines in the blue suitable for cross-correlation analysis. The principle absorption features in this region of the spectrum are Na_D, Mg_b, H_β, and a multitude of Fe lines.

A complete observation of an object consists of a comparison exposure of a helium-neon-argon lamp in both channels, an exposure of the object on one side and sky on the other side, an equal time exposure with sky and object positions reversed (by moving the telescope), and finally another comparison exposure equal to the first. For a 14.5^m galaxy this full procedure requires 10–40 min, depending on the surface brightness of the source. Observations of sources to be used as templates have an extra comparison exposure inserted between the two object exposures. After each night's observing, we take incandescent lamp exposures for 6–8 hr at a count rate comparable to the count rate of the galaxies ($<$ 3000/s). These data are later used to normalize out small-scale sensitivity variations in the first photocathode and pattern noise arising in the electronic processor. A discussion of the pattern noise problem is given in Davis and Latham (1979).

The photon counting nature of the detector enables us to examine the spectrum in real time as it accumulates in our minicomputer. There is a graphic display terminal at the telescope and we routinely do simple real-time processing of the data to normalize out pattern noise, to subtract sky contributions to the data, to add files, etc. We integrate with the source on the left until lines such as Mg_b and Na_D are well defined, and then take an equal time exposure with the source on the right. In the case of low-surface brightness galaxies, we pray we will detect Hα emission. We are presently modifying our operating system so that correlation analysis can be done in real time providing the observer with an objective estimate of the velocity and dispersion error. Typical observations accumulate 200 sky subtracted counts/pixel or roughly 300,000 counts in total. We can observe roughly 20 galaxies per clear night; therefore the observation of 1000 galaxies will require 50 clear nights of observing. We plan to complete the 14.5^m sample in the Spring of 1980. In the Spring of 1978, we observed roughly 350 galaxies in this sample.

Analysis of the absorption line data proceeds via cross correlation as described in the following sections. Section III deals with the mathematical justification for the correlation analysis and can be skipped on a first reading. The subsequent sections provide examples of the analysis procedure.

III. THEORY OF CORRELATION ANALYSIS

a) *Introduction*

Let $g(n)$ be the spectrum of a galaxy whose redshift and velocity dispersion are to be found and let $t(n)$ be a template spectrum of zero redshift and instrumentally broadened stellar-line profiles. These spectra are discretely sampled into N bins, labelled by bin number n; the relationship between wavelength and bin number is

$$n = A \ln\lambda + B \tag{1}$$

Because the spectra are binned linearly with $\ln\lambda$ a velocity redshift is a uniform linear shift. The spectra are assumed periodic with period N for the purposes of discrete Fourier transforms and correlation functions derived herefrom. In addition, the spectra are continuum subtracted, endmasked, and filtered to remove low-frequency spectral variations which arise from continuum variations and non-uniform photocathode sensitivity, and high-frequency noise components beyond the resolution. The endmasking assures that mismatch between the ends of the spectrum is removed.

Let $G(k)$ and $T(k)$ be the corresponding discrete Fourier transforms defined by

$$G(k) = \sum_n g\,(n) \exp(-2\pi i n k/N) \tag{2}$$

(all sums run from 0 to N − 1). Let σ_g and σ_t be the rms of the spectra:

$$\sigma_g^2 = \frac{1}{N} \sum_n g\,(n)^2. \tag{3}$$

Finally, define the normalized cross-correlation function $c(n)$:

$$c(n) \equiv g \times t(n) = \frac{1}{N\sigma_g\sigma_t} \sum_m g\,(m)\, t(m-n). \tag{4}$$

Note that the cross-correlation product, ×, is not commutative. It is defined so that if, for example, $g(n)$ is exactly the same as $t(n)$ but shifted by d units, $gxt(n)$ will have a peak of 1 at $n = d$. Fourier transforming the formula for $c(n)$ gives:

$$C(k) = \frac{1}{N\sigma_g\sigma_t} G(k)\, T^*(k), \tag{5}$$

where * indicates complex conjugation.

b) *Estimation of Redshift and Velocity Dispersion*

Suppose now that $g(n)$ is some multiple α of $t(n)$, but shifted by an amount δ and broadened by convolution with a symmetric function $b(n)$. That is

$$g(n) \cong \alpha\, t * b(n-\delta). \tag{6}$$

Here * means convolution product and is not to be confused with complex conjugation. In order to estimate the parameters α and δ, consider minimizing the expression

$$\chi^2\,(\alpha, \delta; b) = \sum_n [\alpha\, t * b(n-\delta) - g(n)]^2. \tag{7}$$

Notice that this uniform weighting actually weights quadratically any deviations from zero, which means that strong lines are much more strongly weighted than weak lines. This is a desirable feature because strong lines are delineated with better signal-to-noise, and the quadratic dependence can be justified as follows. The noise in a line is largely from the Poisson statistics of the subtracted continuum and so is essentially independent of line strength. The error in determining the position of a given line is inversely proportional to the signal-to-noise or just the line strength. A proper weighted average of many lines would then weight each line by the square of its strength, as in Eq. (7). We have experimented with different weighting schemes by applying nonlinear transformations to spectra (logarithmic, square root) and then correlating; we find that indeed this quadratic weighting is the best.

Equation (7) for χ^2, when rephrased in Fourier space, becomes

$$\chi^2(\alpha, \delta; b) = \sum_k \left[\alpha\, T(k)\, B(k) \exp\left(-\frac{2\pi i k \delta}{N}\right) - G(k)\right]^2. \tag{8}$$

This is equivalent to the formula used by SSBS when $b(n)$ is taken to be a Gaussian except that the weights given each Fourier component are different in the SSBS method, which analyses the Fourier quotient $G(k)/T(k)$. At first it is not apparent that there is any weighting scheme used in our determination of α, δ, and b from minimization of χ^2, but in fact there are two. The first is that each spectrum is filtered prior to analysis by multiplication of its Fourier transform by a bandpass function. Thus some Fourier components are removed entirely and others are weighted according to the value of the bandpass. The second set of weights arises when the parameters are determined from a quadratic least-squares fit to the correlation peak; the fit is affected more by some Fourier components than others.

Rewrite χ^2 as

$$\chi^2(\alpha, \delta; b) = \alpha^2 \sum t * b(n-\delta)^2 - 2\alpha \sum t * b(n-\delta)\, g(n) + \sum g(n)^2$$
$$= \alpha^2 N \sigma_{t*b}^2 - 2\alpha N \sigma_g \sigma_t\, c{*}b(\delta) + N\sigma_g^2,$$

where

$$\sigma_{t*b}^2 = \frac{1}{N} \sum (t*b)^2.$$

Minimization with respect to α yields

$$0 = \frac{\partial \chi^2}{\partial \alpha} = 2N\,[\alpha\, \sigma_{t*b}^2 - \sigma_g \sigma_t\, c{*}b(\delta)],$$

$$\alpha_{\min} = \frac{\sigma_g \sigma_t}{\sigma^2_{t*b}} c*b(\delta). \quad (9)$$

Substituting this value gives

$$\chi^2(\alpha_{\min}, \delta; b) = N\sigma_g^2 \left(1 - \frac{\sigma_t^2}{\sigma^2_{t*b}} [c*b\ (\delta)]^2\right). \quad (10)$$

At this point it is clear that minimization of χ^2 is equivalent to maximization of

$$\frac{1}{\sigma_{t*b}} c*b(\delta).$$

Further progress requires more specific assumptions about t, c, and b, however. First require that $b(n)$ be a Gaussian of dispersion σ:

$$b(n) = \frac{1}{\sqrt{2\pi}\sigma} \exp\frac{-n^2}{2\sigma^2}, \quad (11)$$

$$B(k) = \exp\left(-\frac{(2\pi\sigma k)^2}{2N^2}\right). \quad (12)$$

Assume also that the largest peak in $c(n)$ is approximately Gaussian shaped, centered at δ and with dispersion μ:

$$c(n) \cong c(\delta) \exp\left(-\frac{(n-\delta)^2}{2\mu^2}\right), \quad (13)$$

$$C(k) = \sqrt{2\pi}\,\mu\, c(\delta) \exp\left(-\frac{(2\pi\mu k)^2}{2N^2}\right) \exp\left(-\frac{2\pi i \delta k}{N}\right). \quad (14)$$

Finally, suppose that $t(n)$ has a Fourier transform that is approximately Gaussian in amplitude but with unconstrained phases and a dispersion appropriate for a typical feature in $t(n)$ to have width τ:

$$|T(k)| = \sigma_t \left(\frac{2\pi N\tau}{\sqrt{\pi}}\right)^{1/2} \exp\left(-\frac{(2\pi\tau k)^2}{2N^2}\right). \quad (15)$$

Then expressions for σ_{t*b} and $c*b\ (\delta)$ are easily obtained using the approximate formula

$$\sum_n \exp(-n^2/\sigma^2) \simeq \sqrt{\pi}\sigma, \quad (16)$$

$$\sigma^2_{t*b} = \frac{1}{N}\sum t*b(n)^2 = \frac{1}{N^2}\sum |T(k)\,B(k)|^2$$

$$= \frac{1}{N^2}\sigma_t^2 \frac{2\pi N\tau}{\sqrt{\pi}} \sum \exp\left[-\left(\frac{2\pi\tau k}{N}\right)^2\right] \exp\left[-\left(\frac{2\pi\sigma k}{N}\right)^2\right]$$

$$= \sigma_t^2 \frac{\tau}{\sqrt{\sigma^2+\tau^2}}, \quad (17a)$$

$$c*b(\delta) = \frac{1}{N}\sum_k C(k)\,B(k) \exp\left(\frac{2\pi i\delta k}{N}\right)$$

$$= \frac{1}{N}\sqrt{2\pi}\mu\, c(\delta) \sum \exp\left(-\frac{(2\pi\mu k)^2}{2N^2}\right) \times \exp\left(-\frac{(2\pi\sigma k)^2}{2N^2}\right)$$

$$= c(\delta)\frac{\mu}{\sqrt{\mu^2+\sigma^2}}. \quad (17b)$$

Although, of course, $c(n)$ and $t(n)$ will not truly be Gaussian, nevertheless these are found to be suitable approximations. The parameter δ is chosen as the center of the largest peak in $c(n)$ and as long as the peak is symmetric, this will also be the center of the largest peak in

$$\frac{1}{\sigma_{t*b}} c*b\ (\delta).$$

The proper value for σ can be found by maximizing

$$\frac{1}{\sigma_{t*b}} c*b(\delta).$$

$$0 = \frac{\partial}{\partial\sigma}\frac{1}{\sigma_{t*b}} c*b(\delta) = 2\sigma\frac{\partial}{\partial\sigma^2}\frac{1}{\sigma_{t*b}} c*b(\delta)$$

$$= \frac{\mu\sigma}{2}\frac{c(\delta)}{\sigma_t\sqrt{\tau}}(\sigma^2+\tau^2)^{-3/4}(\sigma^2+\mu^2)^{-3/2} \times [\sigma^2+\mu^2-2(\sigma^2+\tau^2)].$$

Thus the value of σ that minimizes χ^2 is

$$\sigma^2 = \mu^2 - 2\tau^2. \quad (18)$$

This is perfectly reasonable; the width of the cross-correlation peak is an average of the widths of galaxy lines quadratically added to widths of template lines, and is therefore the quadratic sum of two stellar widths and the velocity broadening width. If the instrumental broadening is the same for both the template and galaxy spectra, it is neatly cancelled in the subtraction.

To summarize, the galaxy spectrum is correlated against the template spectrum and the resultant peak is fit by a smooth symmetric function (we use a quadratic polynomial). The central height of this fit determines α, the center is δ, and the width in conjunction with the width of the template provides σ. Practical difficulties will be discussed later in the data-reduction section.

c) *Error Analysis*

In general $c(x)$ will have many peaks, which in cases of low signal-to-noise can distort or surpass the true redshift peak. A peak is selected, and although the data are indiscretely binned, fitting a smooth curve provides accurate values δ, h, and w for the position, height and full-width half-maximum of the peak. This section will investigate error estimates for these parameters as well as confidence levels on the choice of peak.

i) *Error Analysis—Redshift Error*

Errors derived from counting statistics are insufficient because usually the largest error arises from mismatches between galaxy and template spectra. For high quality data the χ^2 per degree of freedom, when properly normalized for the Poisson counting statistics, can exceed 10. In order to account for this mismatch, regard $c(x)$ as the sum of a perfect correlation function of broadened,

shifted template against template plus a remainder function. The first part is supposed to provide a peak of exactly the correct width in the proper position, and the addition of the second part can distort this peak. Specifically suppose

$$c(n) = h_0\, t{*}b \times t(n - n_0) + a(n). \qquad (19)$$

h_0 is used because the observed height h may differ if $a(n_0) \neq 0$.

In order that all the height dependence is contained in h_0, $b(n)$ is assumed to be normalized so that

$$t{*}b \times t(0) = 1.$$

$c(n)$ is known and $a(n)$ is a random function to be determined. Notice that $t{*}b \times t(n - n_0)$ is symmetric about $n = n_0$ because $b(n)$ is symmetric and any autocorrelation function is symmetric about 0. Thus the antisymmetric part of $c(n)$ about $n = n_0$ is the antisymmetric part of $a(n)$ about $n = n_0$:

$$\tfrac{1}{2}[c(n + n_0) - c(-n + n_0)] = \tfrac{1}{2}\,[a(n + n_0) - a(-n + n_0)].$$

In practice n_0 is not precisely known but approximated by δ, and because this procedure gives no information about the symmetric part of $a(n)$, only the rms of the antisymmetric part of $c(n)$ about δ is computed. If it is further assumed that $a(n)$ has symmetric and antisymmetric components uncorrelated but of equal rms, then rms of $a(n)$ is $\sqrt{2}$ times the rms of its antisymmetric component:

$$\sigma_a^2 \cong \frac{1}{N}\frac{1}{2}\sum [c(n + \delta) - c(-n + \delta)]^2. \qquad (20)$$

Figure 10 discussed below gives an example of $a(n)$. Now, regard the error $\varepsilon = \delta - n_0$ as arising from a positive or negative peak in $a(n)$ near enough to n_0 to slew the peak in $t{*}b \times t(n - n_0)$. Estimate this shift by approximating these peaks as parabolas:

$$h_0\, t{*}b \times t\,(n - n_0) \cong h_0\left(1 - \frac{2}{w_0^2}(n - n_0)^2\right).$$

In general, if two parabolas centered at n_0 and n_1 with second derivatives $d_0(n_0)$ and $d_1(n_1)$ are added together, the resulting parabola has a peak at

$$\delta = \frac{n_0 d_0 + n_1 d_1}{d_0 + d_1}. \qquad (21)$$

In this case the true peak has $d_0 = -4h/w^2$, whereas a typical peak in $a(n)$ has height $\sqrt{2}\,\sigma_a$ and width w if we assume the power spectrum of $a(n)$ to be similar to that of $t{*}b \times t$. To first order it is the nearest peak to n_0 in $a(n)$ that is important, and its mean distance from n_0 can be estimated as

$$|n_1 - n_0| \cong \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{N}{2B}, \qquad (22)$$

where B is the highest wave number where the Fourier transform of $c(n)$ has appreciable amplitude (e.g., the half-maximum point). This arises because a highest frequency of B provides $2B$ positive and negative peaks in N bins or approximately $N/2B$ bins between peaks and the mean distance from the nearest is

$$\frac{1}{4}\frac{N}{2B}.$$

Define r as the ratio of the height of the true peak to the average peak in $a(x)$:

$$r = \frac{h}{\sqrt{2}\sigma_a}. \qquad (23)$$

Then the mean error is given by

$$\varepsilon = \delta - n_0 = \left[n_0\left(-\frac{4h}{w^2}\right) + \frac{1}{4}\frac{N}{2B}\left(-\frac{4\sqrt{2}\sigma_a}{w^2}\right) + n_0\left(-\frac{4\sqrt{2}\sigma_a}{w^2}\right)\right] \Big/ \left(-\frac{4h}{w^2} - \frac{4\sqrt{2}\,\sigma_a}{w^2}\right) - n_0,$$

$$\varepsilon = \frac{1}{4}\frac{N}{2B}\frac{1}{1 + r}. \qquad (24)$$

In practice the precise coefficient multiplying $1/(1 + r)$ is adjusted so that the predicted error fits external errors with respect to independently known velocities; this estimate is found to agree closely with our effective $N/8B$ (see Sec. VI).

ii) Error Analysis—Velocity Dispersion Error

The FWHM w of the observed peak can differ from the true FWHM w_0 from two causes: a spurious peak can fall off center and distort the true peak or it can fall on center but with a width other than w_0 and so change the overall width. The latter effect is negligible compared to the former in practical cases. Carrying through a tedious analysis exactly like the above leads to the result that the expected mean error Δw is given by an expression identical to Eq. (24) (Tonry 1979),

$$\Delta w \cong \varepsilon. \qquad (25)$$

Before converting this measure to an uncertainty in σ, the intrinsic Gaussian dispersion of the object, we must define a relationship between w, the FWHM of a second-order polynomial and γ, the width of a Gaussian profile. Let

$$\gamma = f(w). \qquad (26)$$

Because we fit a polynomial to 11 points around the peak of the profile, and because our correlations are bandpass filtered and have negative tails, $f(w)$ is not quite a linear function. $f(w)$ is plotted in Fig. 1 for our choice of filter, which is described in Sec. IV. This plot was empirically generated by fitting polynomials to filtered Gaussians.

Combining Eqs. (26) and (18), we have an expression for the rms error $\Delta\sigma$ of a dispersion:

$$\Delta\gamma = \frac{\partial f}{\partial w}(w)\,\Delta w, \qquad (27)$$

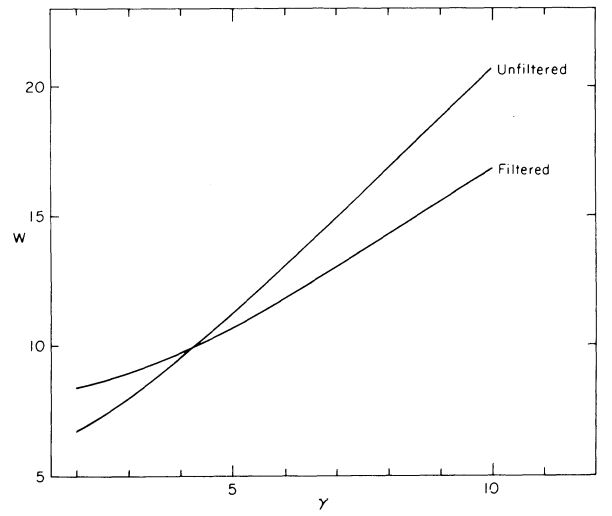

FIG. 1. $f(w)$ relates the measured FWHM w to the Gaussian dispersion. The curve labeled "filtered" results when using the bandpass filter described in the text.

$$(\Delta\sigma)^2 = 2(\mu^2\Delta\mu^2 + 4\tau^2\Delta\tau^2 + \sigma^4)^{1/2} - 2\sigma^2. \quad (28)$$

Assuming the errors $\Delta\mu$ and $\Delta\tau$ to be uncorrelated, Eq. 28 is valid even if $\Delta\sigma > \sigma$. When $\sigma > \tau$ the error in dispersion is comparable to the error in velocity, but if $\sigma < \tau$ the error in sigma is much larger than the velocity error. This is expected because in the latter case, we subtract two large numbers seeking a small difference. Results discussed in Sec. V bear out this expectation.

iii) Error Analysis—Confidence Levels

In cases of poor signal-to-noise, it can become difficult to recognize the true peak of the correlation function because there are other peaks of comparable or even greater height. Generally, the largest is chosen, or correlations against three different templates are made and the peak that shows up best in all three is used. A confidence level can, however, be placed on the choice of peak. Again, using the model described by Eq. (19), the term $ht*b \times t$ will contribute very little to false peaks because the autocorrelation peak at 0 is much larger than any other. $a(n)$ has about $2B$ positive and negative peaks whose heights have roughly a Gaussian distribution of dispersion $\sqrt{2}\sigma_a$. The distribution is Gaussian because $a(n)$ is a sum of approximately B sines of different frequency and random phases so that each peak is built up of a random walk from zero; therefore the dispersion is equal to the mean. The normalized probability distribution of a noise peak of height $h' > 0$ is then given by

$$P(h')\,dh' = \frac{2}{\sqrt{2\pi}}\frac{1}{\sqrt{2}\,\sigma_a}\exp\left(-\frac{h'^2}{4\,\sigma_a^2}\right)dh', \quad (29)$$

$$P(r')\,dr' = \frac{2}{\sqrt{2\pi}}\exp\left(-\frac{r'^2}{2}\right)dr', \quad (30)$$

where r' is defined analogously to r

$$r' = h'/\sqrt{2}\,\sigma_a. \quad (31)$$

In addition, only peaks within a certain region are acceptable, say $-1000 \le cZ \le 20{,}000$ km/sec, which means that only a fraction f of the total number B of positive peaks are considered.

If the true peak has height h and the chosen peak is simply the largest, the probability that the wrong peak has been chosen is the probability that at least one peak of $a(n)$ within the fraction f of admissible redshift has height $\ge h$. Within this fraction there are fB positive peaks so the probability is

$$P(>h) = 1 - \left[1 - \int_h^\infty P(h')\,dh'\right]^{fB}. \quad (32)$$

The confidence level C of the choice of peak is then $1 - P(>h)$, which is rephrased in terms of r as

$$C = \left[1 - \int_r^\infty P(r')\,dr'\right]^{fB}. \quad (33)$$

IV. DATA REDUCTION: DETAILS AND APPLICATION TO M81

We obtain spectra of wide spectral coverage rather than high resolution because our chances of observing some feature suitable for redshift determination are greater. Our pixels are about 1.8 Å in size, our rms resolution is about 2 Å (120 km/sec at 5000 Å); nevertheless, we can often determine redshifts to 20 km/sec and the rms of our error is usually less than 40 km/sec. The data analysis program runs in a Data General minicomputer, taking about two minutes to produce redshift results and errors from initially reduced data, and except in cases of very poor spectra needs no supervision.

It is important to analyze both the emission and absorption lines for redshifts because emission lines can be nonnuclear and have substantially different velocities than the nucleus. In addition, only about 10% of galaxies have very strong emission lines in this spectral region and about 60% do not have sufficient emission to provide a good redshift, even with the convenient digital format of our data. In contrast, 65% have good absorption features and almost all objects provide one type of feature or the other. Thus, although emission lines give an easy and accurate velocity, absorption lines must also be analyzed, and the cross-correlation technique is the most sensitive and accurate method available.

The starting point of this analysis is a sky-subtraced spectrum, a sky spectrum, and a table of comparison-line wavelengths and pixel positions. These are obtained in an initial reduction step that first normalizes the object, sky, and comparison spectra by a high-passed incandescent spectrum; next adds object, sky, and comparison spectra and subtracts the sky spectrum from the object spectrum; and finally goes through the comparison spectrum, fitting parabolas to line peaks and identifying lines with a table of wavelengths of known lines. The main sky lines are identified and their positions and

wavelengths added to the table of comparison lines. We see no systematic differences between the positions of sky lines and comparison lines and inasmuch as there are mercury and oxygen lines in the region 5200–5700 Å where a helium-neon-argon lamp is weak, we use them. A seventh-order polynomial is fit to this table with questionable lines deleted; seventh order is necessary to follow the rapidly increasing distortion of the ends of the spectrum and is not too great to be poorly constrained by the number of lines fit (about 20). The rms residuals of this fit are on the order of 0.2 Å over the full 2600 Å spectral window.

Next the spectrum is scanned for possible emission lines with all candidates fit by a quadratic polynomial and the fit parameters and derived wavelength stored for later analysis. If strong emission lines are present they will dominate the redshift result in the weighted average between emission redshift and correlation redshift, although it is important to recognize that emission regions may have a different velocity than the nucleus which provides the absorption lines for the correlation redshift.

Figure 2(a) is a sky subtracted, normalized spectrum of M81 that has been rebinned into coordinates linear in wavelength. The exposure time was 25 min and so the noise is very small. The pixel to wavelength conversion was achieved by fitting a seventh-order polynomial to 23 lines between pixel 249, wavelength 4471 Å, and pixel 1661, wavelength 6752 Å. The rms residual of the fit was 0.24 Å and the two sky lines used, 5577 and 5890, had residuals of 0.29 and −0.30 Å. Many features are readily apparent. In emission is the [O III] line at 5007, the [N II] lines at 6548 and 6583, Hα, and the [S II] lines at 6717 and 6731. In absorption, the most prominent lines are the Mg I triplet at about 5180, Fe I at 5269, and Na I doublet at 5890. Nearly all the small bumps and wiggles are real features (see Fig. 4).

In preparation for the correlation analysis, a section from 4550 to 6200 Å is mapped to 2048 bins with coordinates linear in logarithmic wavelength

$$n = A\ln\lambda + B.$$

The scale is such that a velocity of about 45 km/sec shifts a feature by one bin. Large sky lines, emission lines, and spikes from image tube hot spots are removed by replacement by a linear interpolation from one side of the feature to the other. A fourth-order least-squares polynomial is fit and subtracted to remove the continuum and the discontinuity between the beginning and end of the spectrum. This subtraction is effective enough that subsequent endmasking of 5% from each end of the spectrum by multiplication with a cosine bell removes all trace of the residual step between the ends of the spectrum and introduces negligible distortion of its own. Figure 2(b) shows the data for M81 after it have been mapped to coordinates linear in logarithmic wavelength, and Fig. 2(c) shows the effects of removal of the fourth-order continuum fit and endmasking.

The final step in preparation of the spectrum is application of the bandpass filter. Our instrumental response behaves like a perfectly sharp spectrum convolved with a Gaussian of dispersion 2 Å. This translates to about 2.7 log wavelength bins, and convolution with a Gaussian of this dispersion in data space is multiplication by a Gaussian of dispersion $N/2\pi \cdot 2.7 \sim 120$ wave numbers in Fourier space. Thus, most information is at

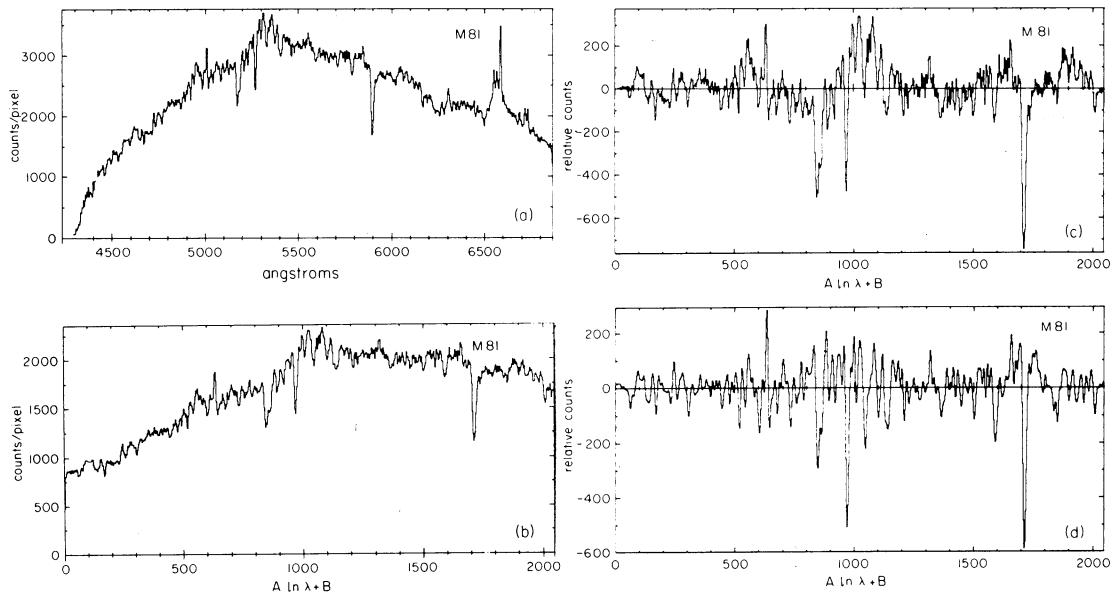

FIG. 2. (a) Spectrum of M81. The ordinate is in accumulated counts per pixel; (b) The result of mapping a segment of the spectrum to ln λ coordinates; (c) A fourth-order least-squares polynomial has been subtracted and the result has been apodized with a cosine bell; (d) The spectrum has now been bandpass filtered and is prepared for correlation analysis.

wave numbers less than 120 and almost everything above wavenumber 240 is noise. Also, wave numbers below 10 contain information about large-scale variations remaining after continuum subtraction, and are ineffective for redshift or dispersion determination anyway. Therefore, the Fourier transform of the spectrum is multiplied by a real bandpass function which starts rising at wave number 10, reaches 1 at 20, starts falling from 1 at 100, and reaches 0 again at 200. The bandpass function is real, so no phase shifts are introduced. Typical features are 9–15 bins broad and have half amplitude points in Fourier space at 60–100 bins, so they are not greatly affected by the filter.

Shown in Fig. 2(d) is the result of bandpass filtering the spectrum shown in Fig. 2(c). Note how both high and low spatial frequency components have been excised. The spectrum is now suitable for the cross correlation.

As a further indication of proper bandwidth to use for cross correlating, we present in Fig. 3 the phase shift versus wave number for the unfiltered correlation of M31 and M81. The slope of the linear regression is a measure of the velocity difference of the two objects; the scatter about this line measures phase error. A band of $+/-\pi$ about the mean slope is delimited by the dashed lines. It is clear that beyond wave number 150, the cross correlation has nearly random phase, and therefore contributes no information.

Shown in Fig. 4 are spectra of M81, M31, M32, the K0 III star, HD12623, and the K2 III star HD52071, all processed as above and then shifted to zero redshift. The objects are all of high signal-to-noise and are all suitable for use as templates. This plot graphically illustrates why the correlation technique is so powerful. Nearly all of the small scale features are seen to be reproduced on each spectrum. Real differences exist of course, such as 5007 emission in M81 and strong H_β absorption in M32, but the major Mg and Na features and the multitude of weak iron lines are quite reproducible, even for objects of different morphological and spectral type.

Some care is required in adjusting the templates to zero redshift, because effects such as off-center guiding on a finite width slit can introduce a redshift error not accounted in the comparison lines. For example, stellar

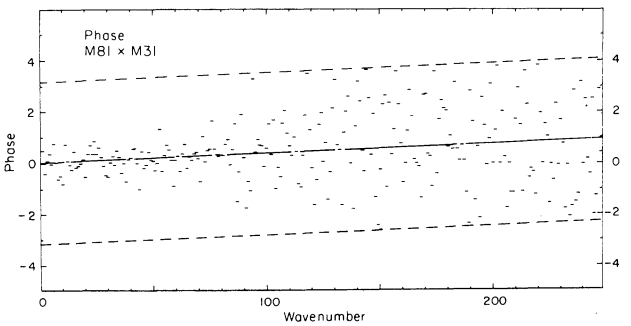

FIG. 3. The measured phase shift of each Fourier coefficient of the correlation function of M81 relative to M32.

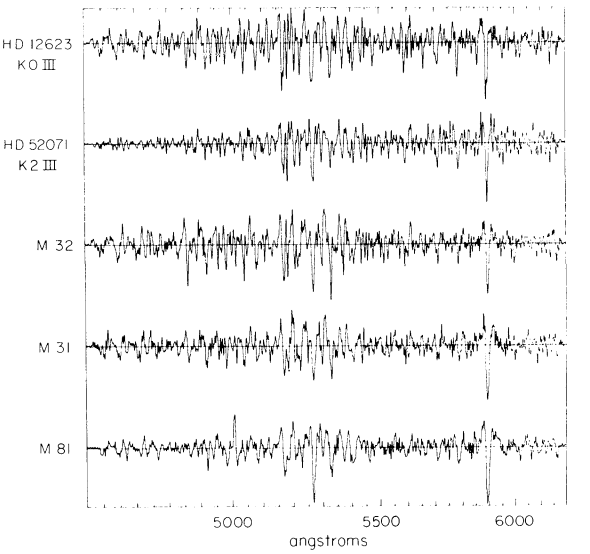

FIG. 4. Spectra of several template quality observations, each shifted to 0 redshift. The effect of internal broadening is apparent in M31 and M81.

spectra generally are further from their true redshift than galactic spectra, which we attribute to a stellar image smaller than the 3.2″ slit width (240 km/sec) spending more time on one side of the slit than the other. Errors of this sort are on the order of 20 km/sec and in order that they not propagate, an independent measure of the redshift is needed. We use the following technique.

A large number (on the order of 50) of good spectra of stars and galaxies with very accurate published velocities are collected, including the templates. Each of these spectra is correlated against all the others and a table of relative velocities and errors is generated. Now if each object is assigned a velocity, another table of relative velocities can be immediately calculated, and the two tables compared. Velocities for each object are obtained by requiring that the sum of the squares of the differences of the two tables weighted by the inverse square of the errors be minimized and that the sum of the assigned velocities equals the sum of the published velocities. The last is what ties the assigned velocities to the published information; relative velocities are unaffected by addition of a constant. This procedure involves inversion of a matrix the size of the number of objects and yields residuals on the order of a few kilometers per second.

The processed object spectrum is correlated against each of the templates by means of fast Fourier transform techniques. A parabola is fit to the largest peak of the correlation function in the range $-1000 < cZ < 20{,}000$ km/s, and the parameters δ (peak position), h (height), and w (FWHM) are obtained.

Using δ as the center position, the antisymmetric part of the correlation function is calculated and σ_a is derived as described in Sec. III. This allows calculation of the parameter $r = h/\sqrt{2}\sigma_a$. Correlations are done for three

templates: the first is an extremely high signal-to-noise sum of galaxy spectra, and the second and third are different stellar spectra. The first template is most effective in producing a recognizable peak in cases of poor signal-to-noise and provides an accurate redshift. Since the template spectrum is continuum subtracted and the depths of lines are on the order of $\frac{1}{5}$ the continuum, the signal-to-noise is $\frac{1}{5}$ that derived from counting statistics and many counts are needed to eliminate Poisson noise domination of the correlation function. For example, 10^5 total counts are usually sufficient to produce a correlation peak against a good template, whereas a good template will require a spectrum of over 10^7 total counts. The second and third templates provide additional redshift estimates, although usually not as good, but their purpose is to allow determination of velocity dispersions.

We correlate each star template against the object spectrum, and the other star template. If the Gaussian instrumental width is τ^2, then the star-star correlation gives a width $2\tau^2$, while the star-object correlations yield widths $2\tau^2 + \sigma^2$. It is better to estimate $2\tau^2$ by correlation of the different stars rather than autocorrelation of a single star, because the perfect match of the spectrum with itself produces an excessively narrow autocorrelation peak. Use of two star templates provides two estimates of σ.

The normalized cross-correlation function of M81 with M32 is shown in Fig. 5(a). Note that there is only 70% correlation between these two objects, and since the Poisson noise is quite small, the remainder is a measure of real spectral mismatch. The antisymmetric part of $c(h)$ about the fit center is shown in Fig. 5(b).

It is instructive to examine the correlation function in Fourier space. For this purpose, unfiltered object and template spectra are used. The amplitude of the Fourier transform initially drops very rapidly, similar to a Gaussian, but has a long tail of high-frequency noise, shown in Fig. 6. Again this confirms our choice of bandpass filter.

Our correlation functions have a half amplitude point in Fourier space of about wave number 50; in data space, they have about 50 positive peaks, so for us, $fB \sim 50/4$. Then, using the formulas from Sec. III

$$\ln(1+Z) = A^{-1}\,\delta, \tag{34}$$

$$A\langle \Delta \ln(1+Z)\rangle = \frac{1}{4}\,\frac{2048}{2.50}\,\frac{1}{1+r} \cong \frac{5}{1+r}, \tag{35}$$

$$C = \left[1 - \sqrt{\frac{2}{\pi}}\int_r^\infty \exp\left(\frac{r'^2}{2}\right) dr\right]^{fB} = \left[1 - \operatorname{erfc}\left(\frac{r}{\sqrt{2}}\right)\right]^{fB} \cong 1 - \frac{1}{4}\times 50\,\operatorname{erfc}\left(\frac{r}{\sqrt{2}}\right). \tag{36}$$

If the correlations against the star templates produce widths w_1 and w_2 with height–noise ratios r_1 and r_2, and the star–star correlation produces width w_0 and ratio r_0, $f(w)$ provides second moment dispersions γ_1, γ_2, and γ_0 yielding, from Sec. III), estimates for σ_v, and $\langle|\Delta\sigma_v|\rangle$:

$$\sigma_v/c = A^{-1}\,(\gamma_1^2 - \gamma_0^2), \tag{37}$$

$$\langle \Delta\sigma_v\rangle^2/c^2 = 2A^{-2}\left[\gamma_1^2\left(\frac{5}{1+r_1}\right)^2\left(\frac{\partial f}{\partial w}(w_1)\right)^2 + \gamma_0^2\left(\frac{5}{1+r_0}\right)^2\left(\frac{\partial f}{\partial w}(w_0)\right)^2 + (\gamma_1^2-\gamma_0^2)^2\right]^{1/2} - 2A^{-2}\,(\gamma_1^2 - \gamma_0^2), \tag{38}$$

and similarly for γ_2.

On the basis of the correlation redshift, a nominal identification of a very strong emission line, or an externally provided estimate, the stored emission line parameters are checked against a table of galaxy emission lines and true lines are identified; each providing an independent redshift value. Probable errors are assigned to these on the basis of line shape, counting statistics of the line strength, and rms residual of the pixel to wave-

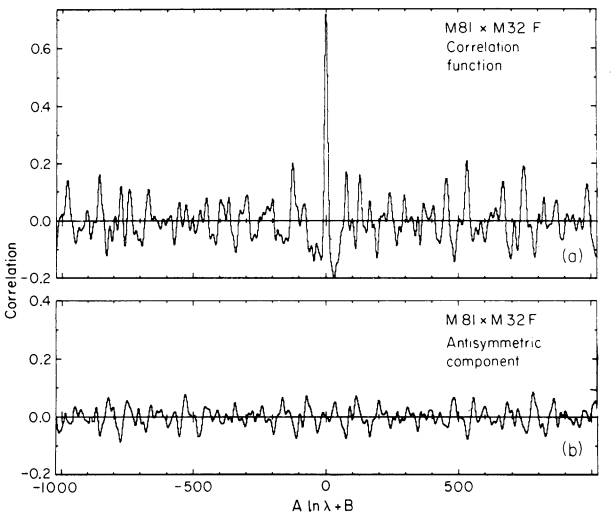

FIG. 5. (a) The cross-correlation function of M81 with M32; (b) The antisymmetric, noise component of this function.

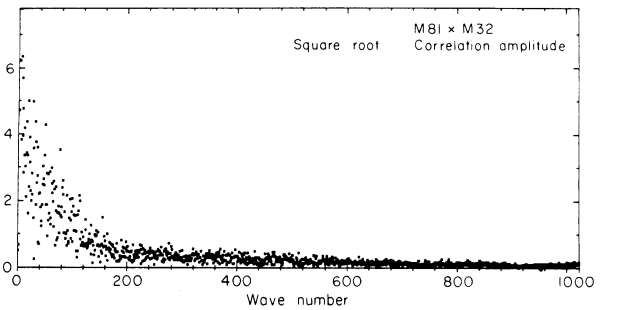

FIG. 6. Amplitude of the Fourier transform of the unfiltered cross-correlation of M81 and M32, showing how the power is concentrated in lower wave numbers.

length conversion, each contribution added quadratically. The probable errors of the correlation redshifts are augmented by adding quadratically a fraction of the rms comparison residual. Weighted averages for emission and correlation redshifts are computed using the inverse squared probable errors as weights. The emission redshifts are treated as independent so the error decreases accordingly, but the three correlation redshifts are not entirely independent, so a weighted average of their errors is used for the overall error.

A major limitation in the accuracy that we can achieve is imposed by flexure of the spectrograph, that is, overall shifts in the spectrum dependent on telescope position, magnetic fields, or other causes. We find that our electrostatic image tubes are very little affected by magnetic fields: with the telescope stationary, movement of the dome, the platform, or other steel objects does not shift the spectrum appreciably. We even clamped magnets to the side of the image-tube package and found very little shift. However, the spectrograph itself is rather flexible, and slewing the telescope 20° can shift the spectrum by as much as 100 km/sec. It is for this reason that we take two equal length comparisons of the start and end of each pair of exposures, we keep our exposures short (< 20 min/side), and we do not slew the telescope while these four spectra are accumulating. A typical shift between beginning and ending comparisons taken in this fashion 30 min apart is 20 km/sec, and if the shift is linear in time the balanced comparison and object exposures are perfectly matched. In order to account for nonlinearity of this shift, all redshift errors have a term of ¼ the mean shift between starting and final comparisons quadratically added.

Finally, a weighted average between emission and correlation redshifts is computed. In cases of significant redshift discrepancies between emission and absorption, we use the absorption redshift if it is of high quality, if it is of low quality, we ignore it in favor of the emission redshift.

The width of the lines in the comparison spectra for the templates and object are compared, and any difference is included as an additional error in the width of the correlation peak which produces an additional error in σ_r.

Table II presents results of this analysis for M81 as well as for other examples discussed below.

V. OTHER EXAMPLES

As examples of realistic data, Figs. 7–13 show the spectrum and correlation function of seven objects correlated against M32. These spectra range from excellent to very poor, the last, IC 101 would be almost useless without the emission line that is probably Hα. The first three are excellent spectra included for comparison with published velocities and velocity dispersions. The fourth N502 is a typical good spectrum, perhaps 50% of our spectra are this good. The fifth, IC 102, is a typical mediocre spectrum, with 75% of program objects as good or better; the sixth, NGC 7786, is poor, with 80% better for correlation but only about 10% as good in emission; and the last, IC 101, is very poor with 15% as bad for correlation and about 50% better in emission. Note how the high degree of symmetry of the cross-correlation function of Figs. 7 diminishes as the signal-to-noise ratio declines in Figs. 8–13.

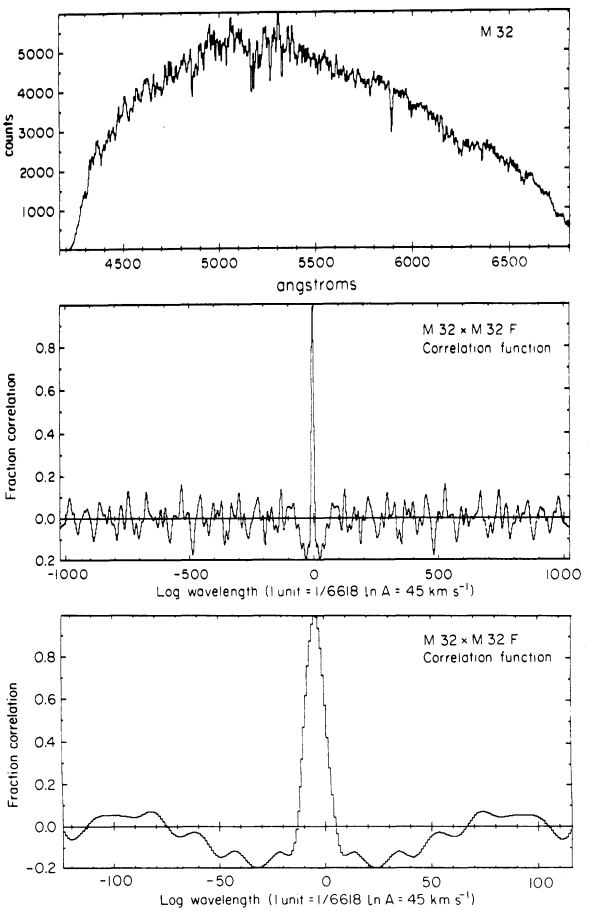

FIG. 7. (a) Spectrum of M32, counts versus wavelength; (b) Auto-correlation function of M32 with itself. Blowup of correlation function near peak region. Note the symmetry of the peak. The template M32 has been adjusted to zero velocity, so the peak is offset from the origin.

Table I summarizes the correlation results for these objects, using templates M32, HD12623, and HD52071. Listed for each are r, the signal-to-noise measure; γ, the RMS correlation width; the confidence of the redshift; and the redshift and dispersion with probable errors.

Table II summarizes the correlation redshifts versus the emission redshifts, and compares measured dispersions against published dispersions. The results for these objects are reasonable. The heliocentric velocities for M32, M31, and M81 agree with published results (not surprising for such high signal-to-noise). In addition, the velocities are self-consistent between correlation templates and between correlation velocities and emission velocities.

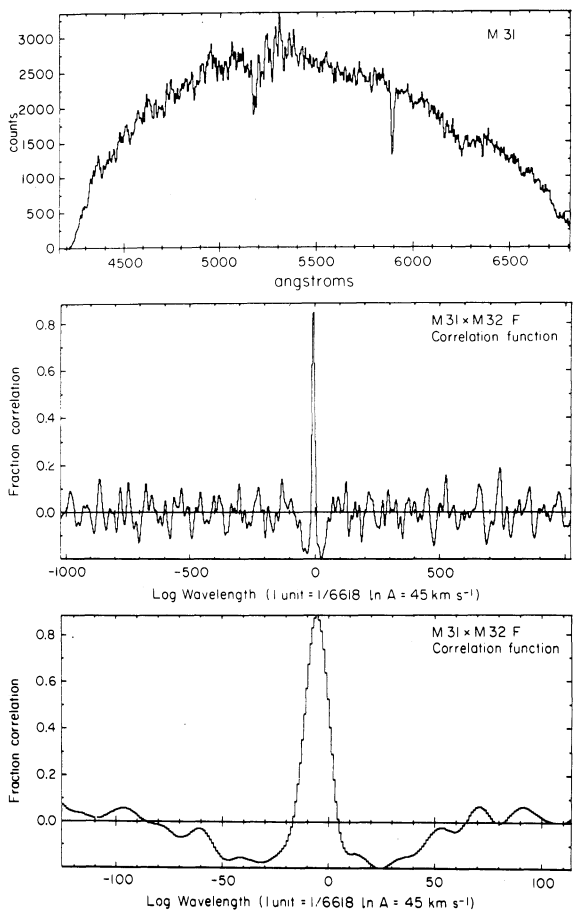

FIG. 8. (a) Spectrum of M31; (b) Cross-correlation M31 × M32; (c) Blowup of peak region. Note the decrease in the symmetry of the peak region.

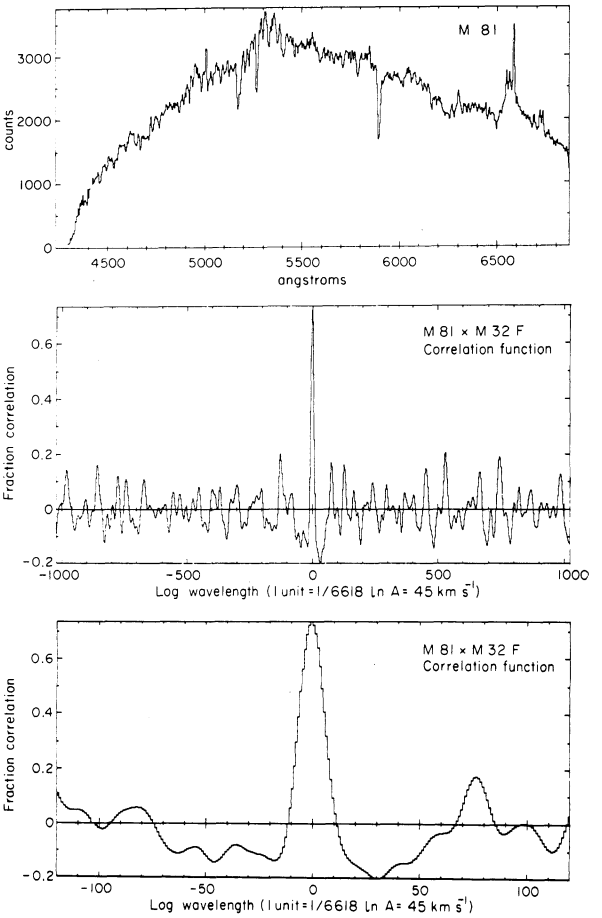

FIG. 9. Same as Fig. 8, for M81.

The velocity dispersions for M32, M31, M81 also are consistent with published results, but the σ values for the other objects should be disregarded since the signal-to-noise is so poor. The errors associated with the velocity dispersions are calculated using Eq. (28). Further study is required to determine if these errors should be scaled by some factor.

Figure 13, the plots of IC101, illustrates the power and weaknesses of this cross-correlation technique. Although the spectrum of IC101 is terrible and would have no discernable individual features if not for the Hα line, choice of the right correlation peak gives a fair redshift. In cases such as this where the signal to noise is low, the correlation function in data space gives a more accurate and less ambiguous result than a simple fit to the phase of the correlation function in Fourier space. The danger is that the wrong peak will be selected, giving a wildly wrong redshift with inappropriate error; the confidence limit here of 60% would be unacceptable without the additional corroboration of the Hα emission.

VI. COMPARISON WITH 21 cm AND OTHER VELOCITIES

The velocities derived from the cross-correlation analysis agree very well within their uncertainties with published velocities. Data from Spring 1978 (approximately 350 galaxies) have been reduced and the overlap with results from other observers examined; it will be published in the redshift catalog of Huchra *et al.* (1979).

Because of the small uncertainty of our velocities, the most sensitive test is comparison with 21 cm data or other highly accurate redshifts. Thirty-two of our observations were of such objects, but two are not used; one because of likely source confusion in the 21 cm measure, and one because of probable error in our choice of correlation peak (the confidence level was about 70%).

The distribution of the quantity x defined as follows

$$x = \frac{v(\text{our}) - v(\text{pub})}{[\varepsilon(\text{our})^2 + \varepsilon(\text{pub})^2]^{1/2}}$$

was examined with the results in Table III. This quantity

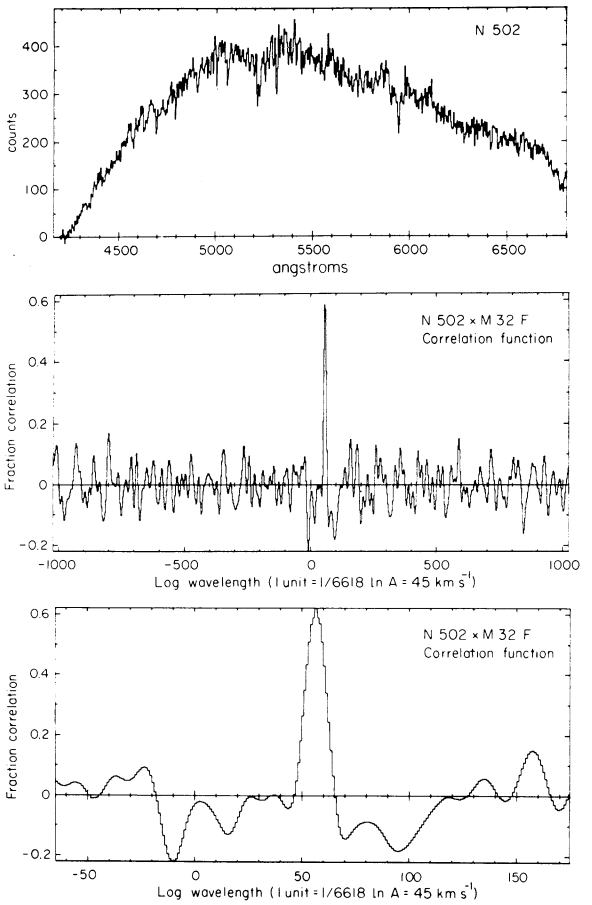

FIG. 10. Same as Fig. 8 for NGC 502.

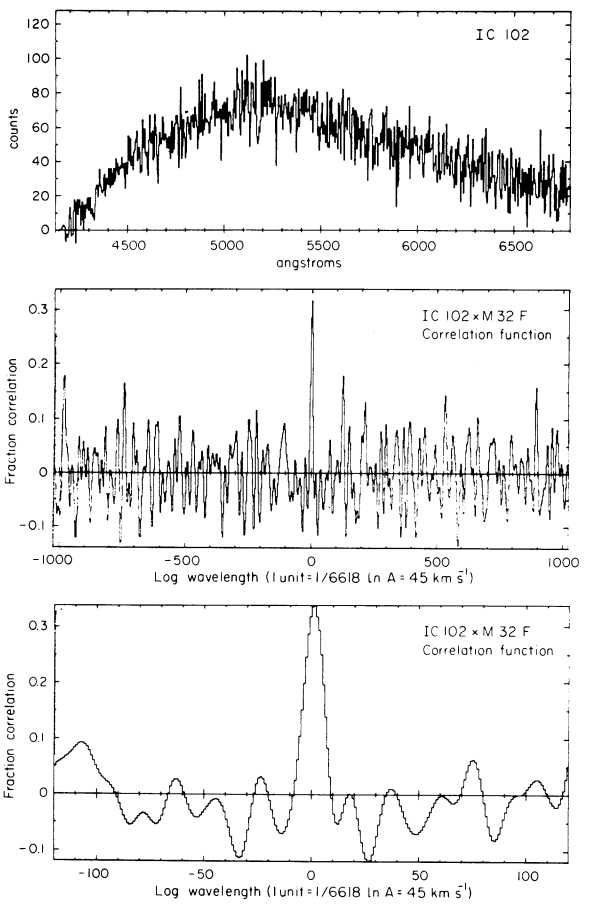

FIG. 11. Same as Fig. 8 for IC102.

should be Gaussian distributed with mean 0 and dispersion 1. The residuals from emission-line velocities, correlation velocities and combined velocities are listed separately for the 21 cm comparison. The emission-line velocities can arise mainly from an emission region with significant orbital velocity that happens to fall in our slit; in fact, half of the deviation of $\langle (x - \bar{x})^2 \rangle^{1/2}$ from a value of 1 was caused by one observation. The distribution of emission-line velocity errors appears to be gaussian with an rms dispersion given by the predicted errors but with an additional extended tail from non-random errors.

The errors in the correlation velocities are subject to the precise constant chosen in Eq. (24), but the natural choice gives excellent agreement with the 21 cm so it is adopted. Thus, the errors assigned to correlation velocities are ultimately tied to external agreement with 21 cm observations. Figure 14 is a plot of correlation error versus the peak significance parameter r for these thirty objects, six different templates and six filters. On the same graph are the curves $\pm 5(1 + r)^{-1}$ which should be at the rms point of the scatter, and indeed they do follow the points except at very low r where incorrect correlation peaks begin to be selected.

Also listed in Table III is the mean predicted error $\langle e \rangle$ which gives an indication of the absolute size of the errors. The value for the correlation velocities is deceptively large because it is inflated by a few very large predicted errors for spectra where the integration was halted when it became apparent that strong emission lines already provided an excellent velocity. This large value of $\langle e \rangle$ disappears when the emission and correlation velocities are combined, but the large value of $\langle (x - \bar{x})^2 \rangle^{1/2}$ persists, again mainly due to the one observation, indicating that care must be exercised in the combining of correlation and emission velocities.

Comparison of our results with those published by Sandage (1978) and those in the Second Reference Catalogue of deVaucouleurs *et al.* (1976) (RC2) which were not already used in the sample above is listed in Table III. Errors for Sandage's velocities were taken to be 2.5 times his internal mean error, as he suggested. In both cases the distribution of x was approximately Gaussian of rms one, but with a tail of anomalous errors. In order to isolate this tail, Table III also gives the moments of the distribution when it is truncated at $|x| < 2.5$.

There are a few significant differences in comparison

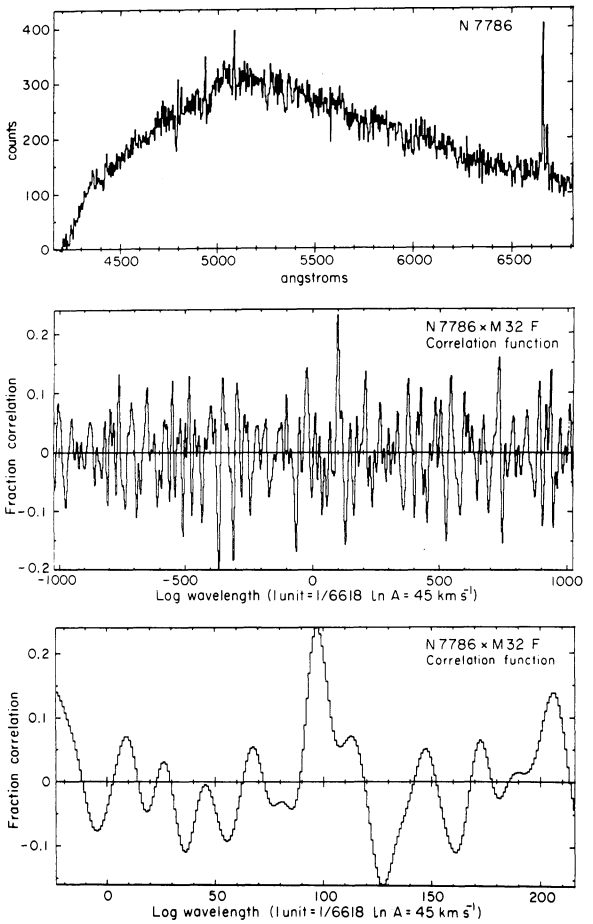

FIG. 12. Same as Fig. 8 for NGC 7786.

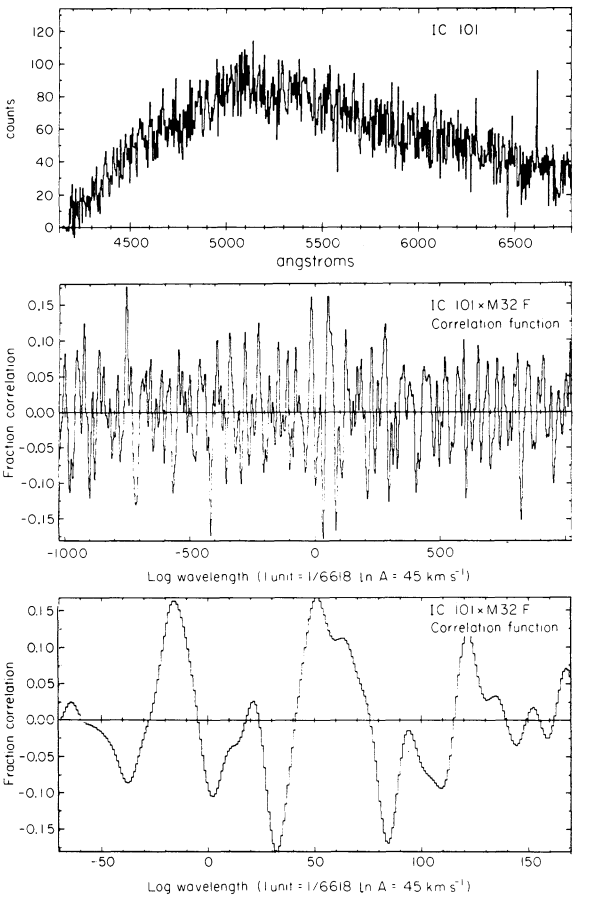

FIG. 13. Same as Fig. 8 for IC101.

with Sandage's results. The four that fell in the truncated tail were cases of disagreement by more than 200 km/sec with spectra that were classified G, F, F, and FP by Sandage. More important is the mean, $\bar{x} = -0.44 \pm 0.13$. It can be reduced to $\bar{x} = -0.08 \pm 0.13$ by removing the +30 km/sec correction used by Sandage to bring his velocity system to the zero point of 21 cm observations, but our overlap with Sandage's results is too small to be sure that a real difference of zero points exists. Our overlap with 21 cm observations is sufficient to preclude a 30 km/sec error in our own zero point.

Anomalies with respect to the RC2 occurred in roughly 20% of all objects in common. One was a case that RC2 noted as an average between two discrepant velocities, another was a situation where the correlation velocity agreed with that listed in RC2 and was in disagreement with the emission velocity; the remaining cases ranged in discrepancy from 100–450 km/sec while quoted error ranged from 20–71 km/sec. When these discrepant cases are removed, $\langle (x - \bar{x})^2 \rangle^{1/2}$ drops to 0.95, as listed in Table III.

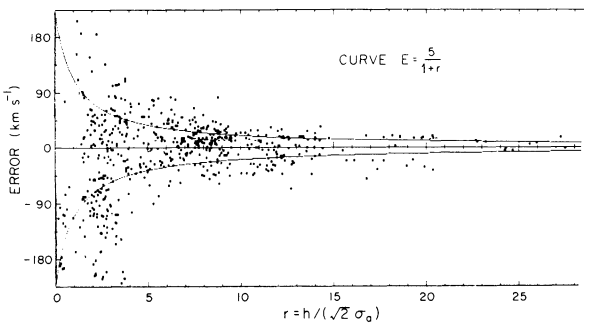

FIG. 14. Scatter plot of error residuals with respect to 21 cm velocities versus signal-to-noise ratio r. The solid curves are the expected rms errors as a function of r.

VII. CONCLUSIONS

The correlation analysis is a clean, unbiased estimator of an object's redshift. As numerous examples show, the method is capable of yielding very high quality redshifts if the signal-to-noise ratio is adequate. The dominant

TABLE I.

Object	Template	$r = h/\sqrt{2}\sigma_a$	Correlation γ	+/−	Confidence	Geocentric (km/sec) cz	+/−	σ_v	+/−
M32	M32	∞	4.54	± 0	1.00	—		—	
M32	HD12623	16.0	4.89	± 0.27	1.00	−215	± 13	71	± 57
M32	HD52021	12.1	4.80	± 0.36	1.00	−210	± 17	57	± 99
M31	M32	23.7	5.68	± 0.17	1.00	−288	± 9	133	± 14
M31	HD12623	12.9	5.92	± 0.29	1.00	−319	± 16	167	± 15
M31	HD52071	12.5	6.08	± 0.30	1.00	−310	± 17	178	± 14
M81	M32	11.5	6.27	± 0.32	1.00	−52	± 18	179	± 17
M81	HD12623	13.5	6.45	± 0.27	1.00	−42	± 16	203	± 11
M81	HD52071	14.8	6.32	± 0.25	1.00	−50	± 14	195	± 10
N502	M32	8.5	4.91	± 0.48	1.00	2537	± 24	31	± 208
N502	HD12623	10.0	4.83	± 0.42	1.00	2555	± 21	62	± 107
N502	HD52071	9.5	4.75	± 0.45	1.00	2543	± 22	49	± 144
IC102	M32	4.4	4.44	± 0.89	1.00	12	± 42	−89	± 165
IC102	HD12623	3.6	4.57	± 1.01	1.00	11	± 49	62	± 389
IC102	HD52071	3.4	4.83	± 1.00	0.99	71	± 51	−33	± 315
N7786	M32	2.6	4.46	± 1.27	0.90	4405	± 63	−87	± 292
N7786	HD12623	2.9	4.79	± 1.14	0.95	4354	± 58	55	± 390
N7786	HD52071	3.7	5.09	± 0.92	1.00	4375	± 49	95	± 195
IC101	M32	2.2	6.82	± 1.26	0.67	2298	± 72	217	± 145
IC101	HD12623	2.0	6.54	± 1.31	0.59	2311	± 75	210	± 153
IC101	HD52071	2.3	4.61	± 1.35	0.77	2193	± 69	−22	± 557

noise can be systematic wavelength errors generated by imperfect guiding through our finite width slit. This will set a limit with the present dispersion at roughly 0.2 Å, or 10–15 km/sec, which has been included in our error estimate and is seen to be a major factor in objects such as M81. However, the correlation process is not magical, as is also illustrated in the figures. If there are no well-defined features in the spectrum, the correlation analysis will produce a poor or meaningless redshift.

Use of the antisymmetric component of the cross-correlation function provides a reliable internal measure of error of the correlation peak. Previous correlation analyses have relied either on external measurement of errors, or on minimum χ^2 analysis. The latter method is suspect because the χ^2, when based on Poisson noise only, is always found to be several times larger than the number of degrees of freedom. Our analysis makes no assumptions about the noise spectrum and provides an internal error that is seen to agree with the external error as determined by comparison to previously published data.

Our method of determination of velocity dispersions is conceptually simple and operationally straightforward. There is no need for a nonlinear least-squares analysis, and no complicated weighting algorithm as required in the Fourier quotient method of SSBS. Our results to date are limited but in due course we shall examine all of our data to determine dispersions.

TABLE II. All velocities are heliocentric cz.

Object	Magnitude	Exposure (Min)	Emission	Emission velocity	Correlation peak $\langle r \rangle$	Average emission velocity	Average correlation velocity	published velocity	Average σ	Publ. σ
M32	9.4	24	—	—	14	—	−209 ± 21	−217 ± 9	64 ± 78	83[a]
M31	4.6	20	—	—	16	—	−294 ± 19	−299 ± 3	159 ± 15	173[b]
M81	7.9	26	5007 (OIII)	−78 ± 21	13	−55 ± 15	−34 ± 22	−44 ± 5	192 ± 12	153[b] ± 17
			6548 (NII)	−64 ± 18						
			6563 Hα	−23 ± 18						
			6584 (NII)	−43 ± 11						
			6717 (SII)	−83 ± 18						
			6731 (SII)	−63 ± 18						
N502	13.8	23	—	—	9	—	2554 ± 27	—	47 ± 153	—
IC102	15.6	27	—	—	3.8	—	21 ± 49	—		—
N7786	13.9	40	4861 H_β	4393 ± 35	3.1	4362 ± 19	4362 ± 57	—		—
			5007 (OIII)	4391 ± 34						
			6563 Hα	4343 ± 15						
			6584 (NII)	4378 ± 21						
IC101	15.1	32	6563 Hα	2379 ± 29	2.2	2379 ± 29	2256 ± 73	—	—	

[a] SSBS.
[b] Whitmore, Kirshner, and Schechter.

TABLE III.

Velocity measure	Comparison source	Number of overlaps	km/sec ⟨ε(our)⟩	km/sec ⟨ε(pub)⟩	$\bar{x}$	$\langle(x-\bar{x})^2\rangle^{1/2}$
Correlation	21 cm	30	43	9	-0.06 ± 0.18	1.01
Emission	21 cm	26	29	9	$+0.07 \pm 0.20$	1.56
Both	21 cm	30	24	9	$+0.01 \pm 0.18$	1.47
Both	Sandage	59	28	84	-0.44 ± 0.13	1.32
Both	Sandage, truncated tail	55	28	85	-0.44 ± 0.13	1.02
Both	RC2	37	30	76	$+0.41 \pm 0.16$	2.37
Both	RC2, truncated tail	27	29	89	-0.02 ± 0.19	0.95

It is seen from Table II that good signal-to-noise is required for a decent estimate of the dispersions, but if the signal-to-noise is sufficiently large, we can measure a dispersion as small as 75 km/sec, slightly less than our resolution sigma of 120 km/sec. As expected, the error of the dispersion estimate is much reduced when the dispersion is itself larger than the instrumental resolution. We believe our expression for $\Delta\sigma$ is of the proper form, but more data are required to ascertain whether the coefficient is accurate.

We encourage observers to use this type of analysis procedure because it provides a reliable measurement of error. The redshift catalogue we are compiling incorporates all data published in the literature, and it would be most desirable if these data were of uniform quality, which is of course an unrealistic hope. However, a fair estimate of the redshift error is easily attainable and equally important. As evident from the comparisons with the published literature, the present catalogues are a heterogeneous concatenation. Use of a significance coefficient such as we describe, or a quality specification as used by Sandage (1978), is a highly useful secondary error estimate that can untangle cases of discrepant redshifts.

This research is part of an ongoing team effort at the Center for Astrophysics to complete the redshift survey. Several people have contributed extensively to this effort in instrument development, software development, and observational assistance. We especially acknowledge the participation of Dave Latham, John Huchra, Arthur Goldberg, and Tom Stephenson. Steve Shectman provided hospitality and invaluable assistance when one of us (M.D.) visited his lab to construct the electronic processor.

REFERENCES

Brault, J. M., and White, O. R. (1971). Astron. Astrophys. **13,** 169.
Davis, M., and Geller, M. G. (1976). Astrophys. J. **208,** 13.
Davis, M., Geller, M. G., and Huchra, J. (1978). Astrophys. J. **221,** 1.
Davis, M., and Latham, D. L. (1979). SPIE Proceedings, Tucson.
Davies, R. L. (1978). Ph.D. thesis, Cambridge.
de Vaucouleurs, G., de Vaucouleurs, A., and Corwin, H. (1976). *Second Reference Catalogue of Bright Galaxies* (Univ. of Texas Press).
Fall, S. M. (1975). Mon. Not. R. Astron. Soc. **172,** 23.
Geller, M. G., and Peebles, P. J. E. (1973). Astrophys. J. **184,** 329.
Gregory, S. A., and Thompson, L. P. (1978). Astrophys. J. **216,** 682.
Huchra, J., Davis, M., Thuan, T X., Knapp, G., Tonry, J., Latham, D. W., and Goldberg, A. (1979). In preparation.
Kirshner, R. P., Oemler, H., and Schechter, P. L. (1978). Astron. J. **83,** 1549.
Peebles, P. J. E. (1979). Astrophys. J., in press.
Sandage, A. (1978). Astron. J. **83,** 904.
Sargent, W. L. W., Schechter, P. L., Boksenberg, A., and Shortridge, K. (1977). Astrophys. J. **212,** 326.
Schechter, P. L. (1978). Private communication.
Shectman, S. A., and Hiltner, W. A. (1976). Publ. Astron. Soc. Pac. **88,** 960.
Simkin, S. J. (1974). Astron. Astrophys. **31,** 129.
Tarenghi, M., Tifft, W. G., Chincarini, G., Rood, H. J., and Thompson, L. A. (1979). Astrophys. J., in press.
Tifft, W. G., and Gregory, S. A. (1976). Astrophys. J. **205,** 696.
Tonry, J. T. (1979). Ph.D. Thesis, Harvard University.
Welch, G., Chincarini, G., Rood, H. (1975). Astron. J. **80,** 77.
Whitmore, B., Kirshner, R., Schechter, P. (1979). Astrophys. J., in press.