THE ASTROPHYSICAL JOURNAL, 475:20–28, 1997 January 20

DETECTION OF WEAK LENSING BY A CLUSTER OF GALAXIES AT $z = 0.83$

G. A. LUPPINO
Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, HI 96822; ger@hokupa.ifa.hawaii.edu

AND

NICK KAISER
Canadian Institute for Advanced Research and Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St. George Street, Toronto, M5S 1A7, Canada; kaiser@cita.utoronto.ca

Received 1996 January 30; accepted 1996 July 12

ABSTRACT

We report the detection of weak gravitational lensing of faint, distant background galaxies by the rich, X-ray luminous cluster of galaxies MS 1054−03 at $z = 0.83$. This is the first measurement of weak lensing by a bona fide cluster at such a high redshift. We detect tangential shear at the 5%–10% level over a range of radii $50'' \lesssim r \lesssim 250''$ centered on the optical position of the cluster. Two-dimensional mass reconstruction using galaxies with $21.5 < I < 25.5$ shows a strong peak which coincides with the peak of the smoothed cluster light distribution. Splitting this sample by magnitude (at $I = 23.5$) and color (at $R-I = 0.7$), we find that the brighter and redder subsamples are only very weakly distorted, indicating that the faint blue galaxies (FBGs), which dominate the shear signal, are relatively more distant. The derived cluster mass is quite sensitive to the $N(z)$ for the FBGs. At one extreme, if all the FBGs are at $z_s = 3$, then the mass within a 0.5 h^{-1} Mpc aperture is $(5.9 \pm 1.24) \times 10^{14}\ h^{-1}\ M_\odot$, and the mass-to-light ratio is $M/L_V = 350 \pm 70\ h$ in solar units. For $z_s = 1.5$ the derived mass is $\sim$70% higher and $M/L \simeq 580\ h$. If $N(z)$ follows the no evolution model (in shape) then $M/L \simeq 800\ h$, and if all the FBGs lie at $z_s \lesssim 1$ the required M/L exceeds 1600 h. These data provide clear evidence that large, dense mass concentrations existed at early epochs; that they can be weighed efficiently by weak lensing observations; and that, for a plausible cluster mass, most of the FBGs must lie at high redshift ($z > 1$).

Subject heading: galaxies: clusters: individual (MS 1054−03) — galaxies: distances and redshifts — galaxies: photometry — gravitational lensing

1. INTRODUCTION

The technique of weak gravitational lensing has emerged as a powerful probe both of clusters of galaxies and of the faint blue galaxy (FBG) population. Most weak lensing observations to date have concentrated on low- and intermediate-redshift clusters ($z \sim 0.2$–0.4); for example, A1689 at $z = 0.18$ (Tyson, Valdes, & Wenk 1990; Tyson & Fischer 1995; Kaiser et al. 1996), A2218 at $z = 0.18$ (Squires et al. 1995), MS 1224+24 at $z = 0.33$ (Fahlman et al. 1994), A370 at $z = 0.375$ (Kneib et al. 1994), and Cl 0024+17 at $z = 0.39$ (Bonnet et al. 1994). Clusters in this redshift range are sufficiently far away that they can be imaged efficiently with existing 2048^2 pixel CCD detectors and yet are close enough that the derived mass is little affected by uncertainty in the redshifts of the faint lensed galaxies.

Observing lensing by high-redshift ($z > 0.7$) clusters is more difficult, since for a lens of a given mass the distortion tends to weaken with increasing lens redshift, especially as the lens redshift approaches that of the sources. However, this dependence of the distortion strength on the observer-lens-source geometry potentially provides a powerful constraint on the redshift distribution $N(z)$ of faint galaxies. If the majority of these lie at high redshift ($z > 2$, say), then we should see strong distortion for even the most distant ($z \sim 1$) massive and concentrated clusters, but if the majority of faint galaxies lie at or below $z \sim 1$, then the distortion should fall rapidly as the cluster redshift approaches unity. In this way, one can constrain $N(z)$ at much fainter magnitudes ($I > 24$) than are accessible by spectroscopic surveys, even with the new generation of 8–10 m telescopes.

Smail et al. (1994) tried this experiment by looking for weak lensing in three clusters covering a wide range of redshifts ($z = 0.26$, $z = 0.55$ and $z = 0.89$). A clear lensing signature was seen in the $z = 0.26$ cluster and a somewhat weaker signal in the $z = 0.55$ cluster, but none was seen in the highest redshift cluster, Cl 1603+43 at $z = 0.89$, suggesting that the majority of FBGs with $I < 25$ were at $z \lesssim 1$. However, an alternative interpretation is that Cl 1603+43 is simply not massive enough to produce a measurable shear signal. This is not implausible since this cluster was optically selected (Gunn, Hoessel, & Oke 1986) and has an X-ray luminosity of only $L_X \sim 1 \times 10^{44}$ ergs s^{-1}; (Castander et al. 1994), as compared to the two lower redshift clusters which both have $L_X > 10^{45}$ ergs s^{-1}. Of course, Smail et al. had little to choose from. When they performed their observations, there were no known clusters at $z > 0.7$ with X-ray luminosities comparable to the richest and brightest low-redshift clusters, and the small number of high-z clusters then known were mainly optically detected (e.g., Gunn et al. 1986; Couch et al. 1991). Recently, however, several new, high-redshift clusters have been discovered as the optical counterparts to previously unidentified *Einstein* Extended Medium Sensitivity Survey (EMSS) X-ray sources (Gioia et al. 1990; Gioia & Luppino 1994). The most distant of these, MS 1054−03 at $z = 0.83$, is extremely rich and has an X-ray luminosity an order of magnitude higher than Cl 1603+43 (Luppino & Gioia 1995), suggesting it may be a potent gravitational lens.

In this paper, we report the detection of weak gravitational lensing by MS 1054−03. Our observations and data reduction are outlined in § 2, the cluster properties are described in § 3. In § 4, we apply weak lensing analysis, and in § 5, we discuss the implications of our observations for cosmological structure formation models and for the con-

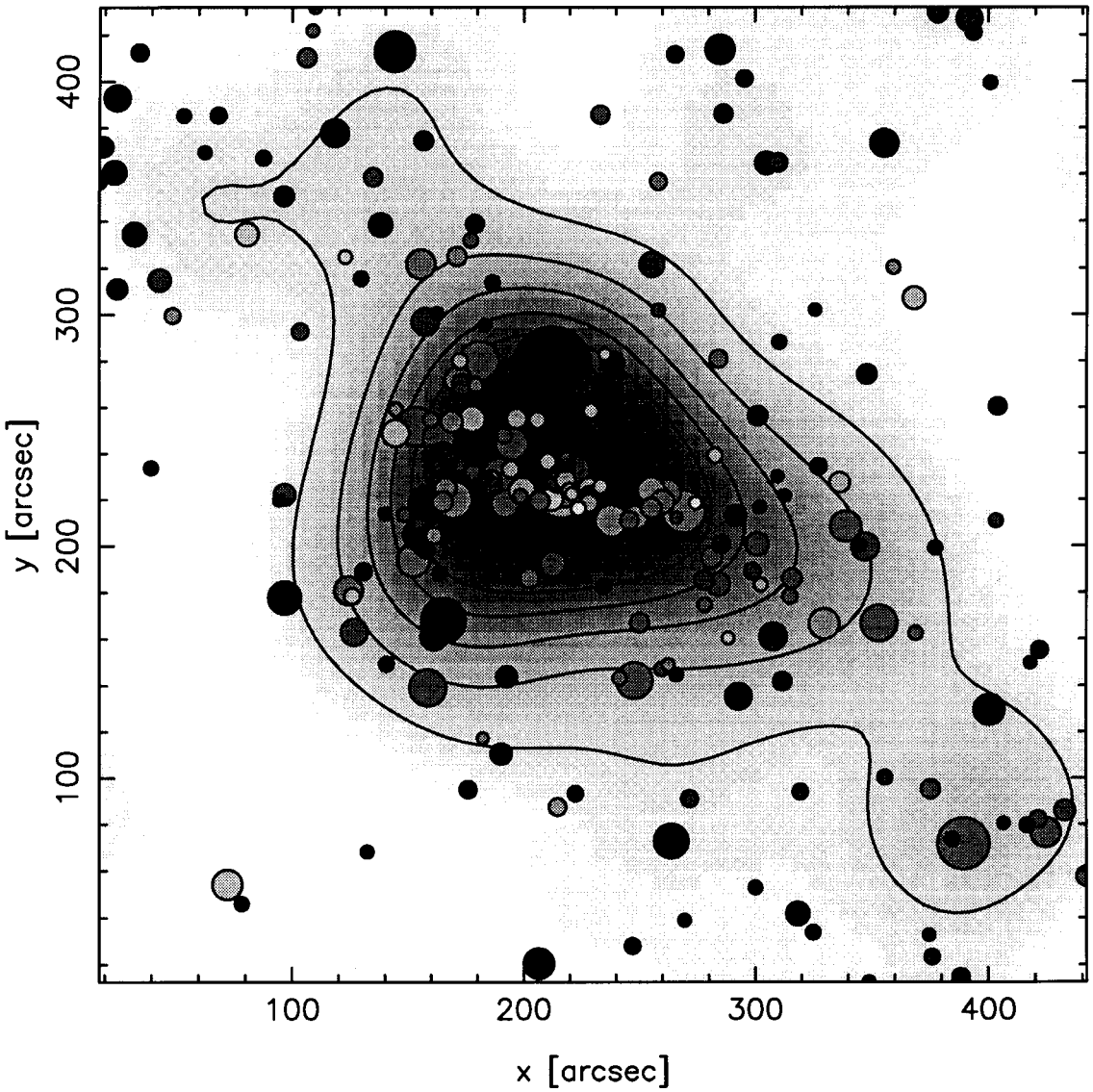

FIG. 2.—Spatial distribution of red galaxies (including the sequence of cluster galaxies with $R-I \simeq 1.5$). The size of each circle is proportional to the brightness of the galaxy (in I), and the shading indicates the color on a scale of $R-I = 1.9$ (*white*) to $R-I = 1.1$ (black). The underlying gray scale is the I-band surface brightness smoothed with a 35″ Gaussian filter.

straining the redshift distribution of the faint background galaxies. Unless otherwise noted, all cosmological quantities are computed using $q_0 = 0.5$ and a Hubble constant parameterized as $h = H_0/100$ km s^{-1} Mpc^{-1}.

2. OBSERVATIONS AND DATA REDUCTION

Optical R- and I-band images of MS 1054−03 were obtained with the UH 2.2 m telescope on the nights of 1993 February 19 and 1994 January 11–13. A thinned Tek 2048^2 CCD was mounted at the f/10 RC focus resulting in a scale of 0.″22 pixel^{-1} and a field of view of 7.′5 × 7.′5 (physical scale 1.86 h^{-1} Mpc at $z = 0.83$). The total exposure times were 7200 s and 21,600 s in R and I, respectively. The individual images in each filter were first debiased and then flattened using a median of all the CCD frames taken in that filter (including the cluster images which made up $\sim\frac{1}{3}$ of the total number of frames). Low spatial frequency residual sky fluctuations were then removed by subtracting a highly smoothed image determined from the troughs of the minima in the images. Registration was performed using ~ 50 moderately bright stars, and the images were then transformed to a common coordinate system (with bilinear interpolation). The stack of transformed images was then summed with cosmic-ray rejection and using appropriate weights (the cosmic-ray rejection being done in such a way as to ensure that the effective psf for the stars was the same as for the faint objects). The seeing in the resulting R and I images was 1.″14 and 0.″97 FWHM, respectively. Photometric calibration was performed using the standard stars of Landolt (1992). The variation in extinction between the I-band images was very small, as was also the case for all but three of the R-band images. The 1 σ surface brightness limits of the summed R and I images are 27.9 mag arcsec^{-2} and 27.8 mag arcsec^{-2}, respectively.

In order to detect the faint objects we used the algorithm of Kaiser, Squires, & Broadhurst (1995, hereafter KSB). This provides a catalog with accurate positions but crude size and magnitude information. We then used this catalog to mask the summed images and thus determine and subtract the small residual positive bias in the images left by the local sky subtraction, and we then applied photometric analysis to obtain refined sizes, magnitudes, etc. The

resulting catalog contained some noise peaks as well as detections of groups of objects. These were removed by limiting the catalog at 5 σ detections and removing abnormally small and large objects. We also rejected a small number of objects with high eccentricity to obtain final catalogs containing $N_I = 2718$ and $N_R = 1822$ objects, corresponding to about 1.7×10^5 and 1.2×10^5 objects per square degree. Nearly all the objects detected in the R band were also detected in I. The I magnitudes were determined using a large aperture $r_{ap} = 3r_g$, where r_g is the smoothing scale at which the object was detected, and typically overestimate total magnitudes by $\lesssim 0.1$ mag.

3. CLUSTER PROPERTIES

MS 1054−03 is an extraordinary object. It is by far the richest and most X-ray luminous high-redshift ($z > 0.7$) cluster known and is among the richest clusters known at any redshift. A true color image centered on the $I = 19.3$ brightest cluster galaxy (BCG) is shown in Figure 1 (Plate 1); the cluster is easily identified as the horizontal swath of red galaxies in the center of the frame. Figure 2 shows the location, I magnitude, and color of all the nonstellar objects with $I < 24.5$ and with colors in the range $1.93 > R-I > 1.1$ which brackets the color of the cluster galaxies. The total magnitude for all of the galaxies contained within a 1′ aperture (physical scale of $\simeq 0.25\ h^{-1}$ Mpc for $q_0 = 0.5$) centered on the brightest cluster galaxy is $I = 16.5$. Converting the observed I-band magnitude to a rest-frame solar luminosity $L_{V\odot}$ using the relation $M_V = I - 5 \log [(1 + z_l)^2 D_l] - 25 + (V-I)_0 - K(z)$ with the K-correction $K(z) = 0.85$, $(V-I)_0 = 1.3$, and $M_{V\odot} = +4.83$ we obtain $L(<0.25\ h^{-1}\ \text{Mpc}) = 1.19 \times 10^{12} L_{V\odot}$, which includes a $\sim$15% contribution from the bright foreground galaxy lying $\sim 1'$ to the north of the cluster center. For a $0.5\ h^{-1}$ Mpc aperture we find $L(<0.5) = 2.0 \times 10^{12} L_{V\odot}$. The number of galaxies with $I < 22$ counted within the same apertures are $N(<0.25) = 49$ and $N(<0.5) = 82$, which represent an excess over the background of about 44 and 67 galaxies, respectively, making this at least a richness class 4 cluster (Bahcall 1981).

Although MS 1054−03 is clearly very X-ray luminous ($L_{0.3-3.5\,\text{keV}} = 9.3 \times 10^{44}\ h_{50}^{-2}$ ergs s^{-1}), the actual X-ray flux is quite low because the cluster is so distant, and consequently little can be said about its X-ray properties at the present time. MS 1054−03 was unresolved in the *Einstein* IPC with only 107.9 ± 12.8 counts in an 18 ks exposure, corresponding to a flux of $f_X = 2.11 \times 10^{-13}$ ergs cm^{-2} s^{-1} (Henry et al. 1992). The flux was converted to a luminosity assuming a 6 keV thermal spectrum and correcting for extended emission as outlined in Gioia & Luppino (1994). An *ASCA* spectrum has recently been obtained, and a preliminary analysis indicates the cluster has a high X-ray temperature (Donahue, private communication). *ROSAT* HRI observations are scheduled.

4. WEAK LENSING ANALYSIS

The weak lensing analysis involves several steps. Object polarizations $e_\alpha = \{I_{11} - I_{22}, 2I_{12}\}/(I_{11} + I_{22})$ were formed from the quadrupole moments

$$I_{ij} = \int d^2\theta W(\theta)\theta_i \theta_j f(\theta),$$

where f is the flux density and $W(\theta)$ is a Gaussian weighting function matched to the size of the galaxy. We then extract a sample of moderately bright stars that have nonzero polarization due to anisotropy of the point spread function, fit a low-order polynomial model for the psf variation across the field,[1] and then correct the galaxy polarizations for all the objects to what they would have been for perfectly circular seeing as described in KSB. These e_α values should now be equal to the random intrinsic values plus a small coherent shift which is proportional to the gravitational shear $\gamma_\alpha = \frac{1}{2}\{\phi_{,11} - \phi_{,22},\ 2\phi_{,12}\}$, where ϕ is related to the dimensionless surface density by $\kappa = \Sigma/\Sigma_{\text{crit}} = \frac{1}{2}\nabla^2\phi$ and where the critical density $\Sigma_{\text{crit}}^{-1} = 4\pi G c^{-2} D_l D_{ls} D_s^{-1} = 4\pi G c^{-2} D_l \beta$, with $\beta \equiv D_{ls}/D_s$ $\{= [1 - D_l(1 + z_l)/D_s(1 + z_s)]$ for $\Omega = 1\}$.

The next step is to calibrate the relation between the polarization and the shear. Previously, this has been done by artificially shearing deep *HST* images to simulate lensing and convolving with a Gaussian seeing disk (KSB). Here we have used a slightly different approach as described in the Appendix. This new approach gives results which agree very well with those from the previous method using *HST* images (KSB), but is more convenient here. This procedure supplies us with a calibration factor for each of the subsamples we analyze below, so for each galaxy we have a fair estimate of the shear $\hat{\gamma}_\alpha = e_\alpha/\langle P_\gamma \rangle$.

First, we define a sample of all faint objects in the I catalog having $I > 21.5$ (2395 objects). No attempt was made to remove stars or cluster galaxies. This faint galaxy sample can be seen in Figure 3 (Plate 2) as ellipses overlaid on the I-band CCD image of the cluster. Figure 4 shows the result of applying two different inversion algorithms to recover the dimensionless surface density $\kappa(\boldsymbol{r})$: the original Kaiser & Squires (1993, hereafter KS93) algorithm and the new, unbiased Squires & Kaiser (1996, hereafter SK96) algorithm. Mass maps generated by either algorithm (see Figs. 4*a* and 4*c*) show strong mass concentrations very close to the peak of the smoothed light map. Also shown are reconstructions using the same spatial distribution, but with random Gaussian shear values with $\langle \gamma_\alpha^2 \rangle^{1/2} = 0.6$ (a value determined from the data as described below). These mass reconstructions have been smoothed to the same 35″ Gaussian filter scale as the light. Figure 5 (Plate 3) shows a contour plot (*white contours*) of the cluster light superimposed on the mass contours (*black contour lines*) overlaid on the I-band CCD image of the cluster field.

While the relation between the shear (essentially the tidal field) and κ is a nonlocal one, there is an explicit local expression for the gradient of the surface density in terms of the gradients of the shear (Kaiser 1995), and one can there-

[1] We used $\sim$60 stars in each passband (64 in I and 55 in R) to measure the variation of the psf across the field. These showed a clear systematic psf anisotropy with some coherent variation across the image. We modeled this as a Taylor series expansion with the coefficients determined by least-squares minimization. We found that a second-order expansion (six coefficients per polarization component) provided an adequate description of the psf variation, with negligible decrease in residuals when we added higher order terms. It is possible that there are very high frequency variations in the psf that we would have failed to model, but these have little or no effect on our mass estimates (which are only sensitive to low spatial frequency terms). Similarly, the dominant component in our model for the psf anisotropy (the zeroth-order term) has no effect on the mass reconstruction. The terms which do couple effectively to the mass estimates are the first- and second-order terms in the expansion (which correspond to gradients and second derivatives of the psf) which are well determined by our model. Even these have a rather minor effect on the mass reconstruction: if we apply no correction whatsoever, the mass estimates only change by $\sim$5%.

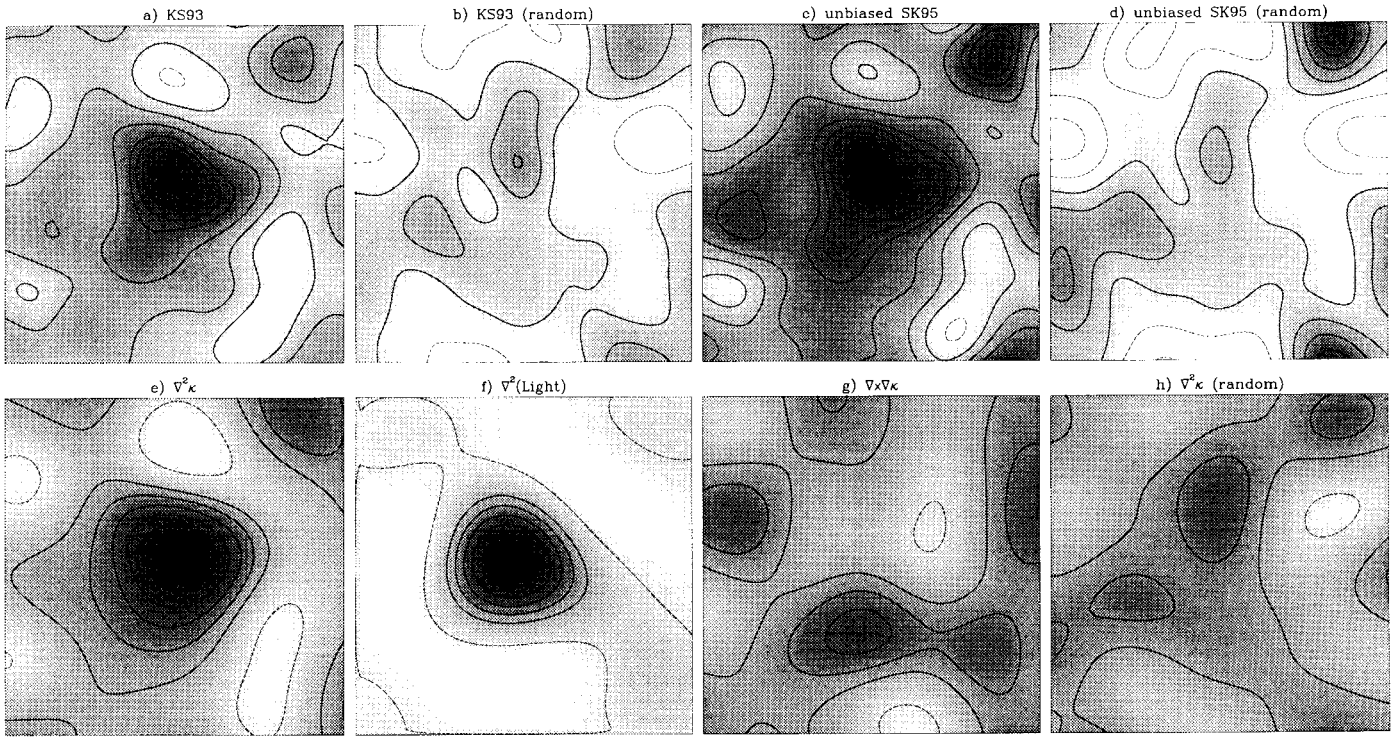

FIG. 4.—The top four panels show the result of two different mass reconstruction algorithms: (*a*) the original KS93 method and (*c*) the new, unbiased "regularized maximum-likelihood" technique of SK96. While the KS93 method is susceptible to a slight negative bias at the edge of the field (Schneider 1995), it appears that in this case any bias that might be present is small. Panels (*b*) and (*d*) are reconstructions using a catalog in which the galaxies were assigned normally distributed random shear values with rms (per component) $\gamma_\alpha = 0.6$, and which indicate the expected level of noise in these reconstructions. The lower four panels contain (*e*) a smoothed image of $\nabla^2\kappa$ (or equivalently κ smoothed with a compensated "Mexican-hat" filter), (*f*) the Laplacian of the surface brightness (scaled to have the same peak value), (*g*) an estimate of $\nabla \times \nabla\kappa$ which should be zero if the shear field is really due to gravity, and (*h*) a realization of the noise produced by our random catalog.

fore determine $\nabla^2\kappa$, the Laplacian of the surface density, from local shear estimates. A smoothed image (filter scale = 70″) of $\nabla^2\kappa$ is shown in Figure 4*e*. The smoothed Laplacian is just the surface density convolved with a particular form of "Mexican-hat" smoothing filter—it is because this filter is "compensated" that the resulting field does not suffer from the slight bias (Schneider 1995) inherent in the KS93 method, and so can be compared directly with the Laplacian of the surface brightness (Fig. 4*f*); clearly these agree in shape and location very well indeed.

An interesting feature of this kind of analysis is that it provides a powerful check on whether the distortion we are detecting is really due to gravitational lensing. If instead of the Laplacian $\nabla \cdot \nabla\kappa$ we calculate the curl of the gradient $\nabla \times \nabla\kappa$, we should then get zero plus fluctuations due to the random noise in the shear estimates. What we are doing here is exploiting the fact that while a general distortion field has two real degrees of freedom, one generated by gravity has only one, and we are projecting out two components of the shear field: one which is excited by gravitational lensing and another which is not. To generate $\nabla \times \nabla\kappa$ rather than $\nabla \cdot \nabla\kappa$, we simply swap the two components of the shear and change the sign of one of them (this is equivalent to rotating each object by 45°). Due to the high symmetry of these operations, one would expect most (but not necessarily all) artificial sources of distortion to excite both modes, and so the smallness of the estimate of $\nabla \times \nabla\kappa$ (visible in Fig. 4*g*) provides a nontrivial check of the reality of the shear field we detect. Finally, the amplitude of the noise fluctuations expected are indicated in the lower right panel of Figure 4, and we see no excess of noise due to artificial sources of image polarization (such as errors in the registration).

To search for variation in the distance to the background galaxies we have split the full $I > 21.5$ sample into subsamples by magnitude (at $I = 23.5$) and color (at $R - I = 0.7$). The mass reconstructions for these four (bright, faint, red, blue) subsamples are shown in Figure 6. The faint and blue reconstructions are very similar. They clearly show the cluster, which now appears elongated in the same sense as the cluster galaxies, and give a somewhat higher peak than for the full sample (although at a similar 5 σ level of significance). The red and bright subsamples, however, show very little sign of the cluster at all—as would be expected if the typical redshift of these objects is less than or of order unity. The smaller shear found for the faint/red galaxies is not due to these galaxies being systematically smaller and consequently rendered more circular by the psf. Since we have determined the shear-polarization calibration factor separately for each subsample, any systematic variation in the size of galaxies between the samples should be correctly accounted for and the difference in shear must reflect a difference in the mean distance to the different subsamples of galaxies.

In addition to the two-dimensional mass reconstruction we have performed "aperture mass densitometry." The statistic

$$\zeta(r) = (1 - r^2/r_{\max}^2)^{-1} \int_r^{r_{\max}} \langle \gamma_T \rangle d\ln r \qquad (1)$$

(Squires et al. 1995; Fahlman et al. 1994) measures $\bar{\kappa}(r)$, the mean surface mass density interior to r, minus the mean

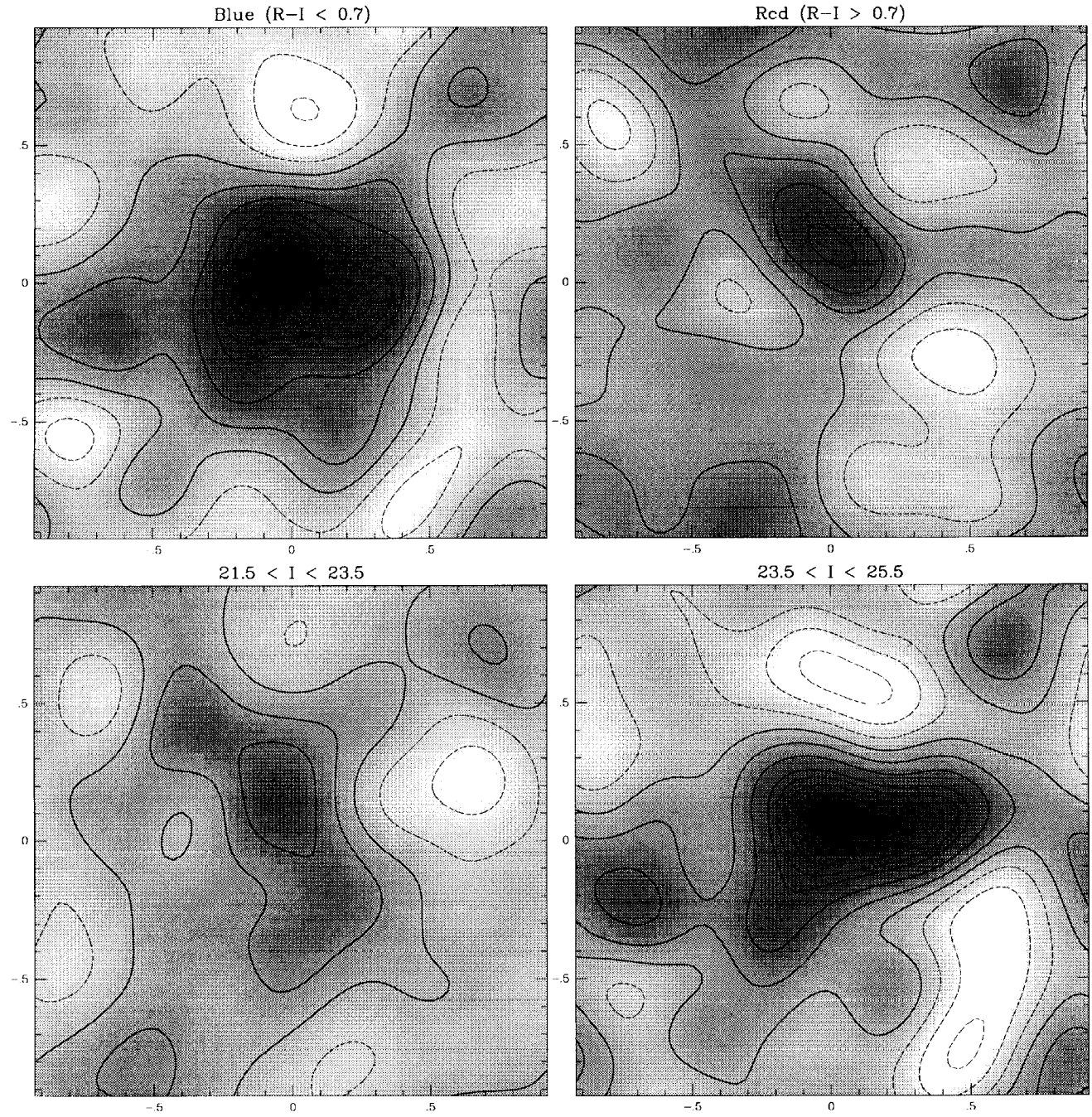

FIG. 6.—Mass reconstruction from the various galaxy subsamples: *upper left*, blue; *upper right*, red; *lower left*, faint-bright galaxies ($21.5 < I < 23.5$); *lower right*, faint-faint galaxies ($23.5 < I < 25.5$). The axes are labeled in units of h^{-1} Mpc. All four mass maps are displayed with the same intensity stretch and contour levels.

surface density in the annulus from r to $r_{\max}$, and therefore provides a lower bound on $\bar{\kappa}$ and hence on the mass within an aperture of radius r. Here the tangential shear is $\langle \gamma_T \rangle = 1/2\pi \int \gamma_T \, d\varphi$, where $\gamma_T = \gamma_1 \cos 2\varphi + \gamma_2 \sin 2\varphi$, and φ is the azimuthal angle with respect to some chosen center (which we have taken to be the peak of the smoothed light image in Fig. 2).

The tangential shear and $\zeta(r)$ are shown for the various subsamples in Figure 7. A coherent tangential shear pattern is clearly seen in the $I > 21.5$ sample over a range of radii from $\sim 50''$ to $\sim 300''$ (although we do not have full azimuthal coverage for $r > 220''$), and the ζ statistic shows that the mean dimensionless surface density rises to $\bar{\kappa} \simeq 0.25$ at $r \simeq 60''$ with a fractional statistical error of about 20%. We calculate the variance in $\gamma_\times \equiv -\gamma_1 \sin 2\varphi + \gamma_2 \cos 2\varphi$. If the shear pattern is circularly symmetric then this should give a fair estimate of the statistical uncertainty in the shear estimates, and the error bars in Figure 7 are based on this estimate. For the $I > 21.5$ sample, for instance, we obtain $\langle \gamma_\times^2 \rangle^{1/2} \simeq 0.6$, which is the value used in the "noise reconstructions" of Figure 4. The γ estimates have uncorrelated statistical uncertainty, whereas the ζ estimates are somewhat correlated (as we have used logarithmically spaced bins in r, each ζ estimate is just a sum of the γ estimates that lie at larger radii; thus ζ estimates at small r tend to have errors which are quite strongly correlated). We should emphasize that because we have taken the spatial origin to be the brightest cluster galaxy, the errors in both γ

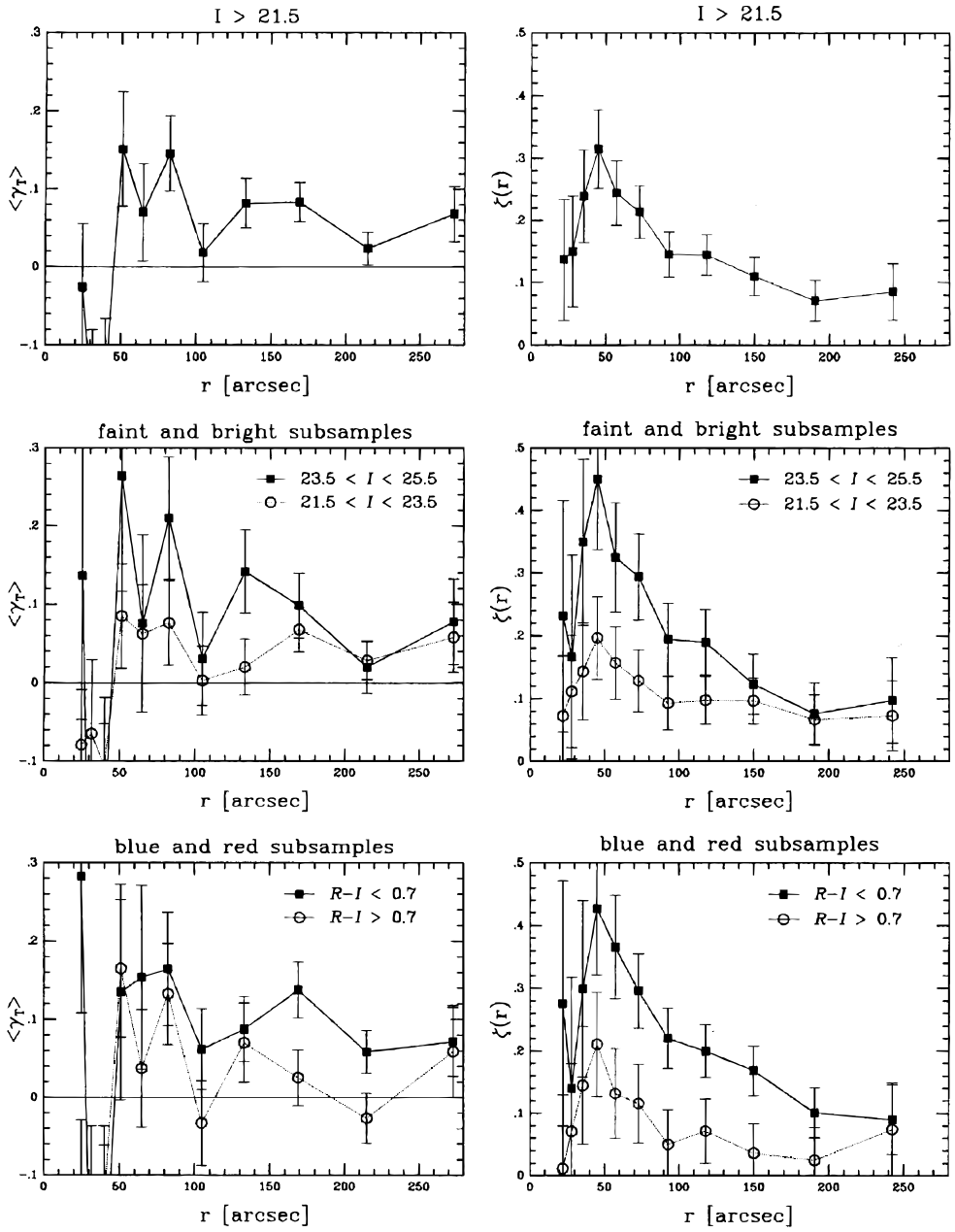

FIG. 7.—Panels on the left show the tangential shear γ_T for the $I > 21.5$ sample (*top*); the faint and bright subsamples are shown as square and circular symbols in the middle panel and the blue (*square*) and red (*circle*) samples are shown in the bottom panel. The right-hand panels show $\zeta(r)$, which provides a lower bound on $\bar{\kappa}(r)$.

and ζ are unbiased, and it is equally likely that we have over- or underestimated the mass.

The lower panels in Figure 7 show graphically how the distortion strength varies with color and magnitude of the background objects. The tangential shear is barely seen in the bright and red subsamples, while for the faint and blue samples, γ_T lies roughly 30% higher than the full $I > 21.5$ sample and gives $\bar{\kappa}(<0.25) \simeq 0.35 \pm 0.07$ and $\bar{\kappa}(<0.5) \simeq 0.20 \pm 0.06$. For the bright and red subsamples, the values are 0.13 ± 0.07 and 0.07 ± 0.05, and this *difference* (in shear values between red and blue or bright and faint subsamples) is significant at the $\simeq 2.2\ \sigma$ level. These values are unlikely to have been significantly affected by cluster contamination, since they only make use of data outside the aperture.

The average physical surface mass density is obtained by multiplying $\bar{\kappa}$ (or ζ) by the critical density, $\Sigma_{\rm crit}$, and a lower limit to the total projected mass within r is then $M(<r) > \pi r^2 \zeta(r) \Sigma_{\rm crit} = c^2 r^2 \zeta / (4 G D_l \beta)$. The big uncertainty here is the value for β, which varies by a factor of ~ 5 from $\beta \sim 0.1$ if all the FBGs are at $z_s \sim 1$ to $\beta \sim 0.5$ if the FBGs are at the maximum plausible redshift of $z_s \sim 3$ (Guhathakurta, Tyson, & Majewski 1990). The critical surface density is $\Sigma_{\rm crit} = 1.95 \times 10^{15} \beta^{-1}\ M_\odot\ h\ {\rm Mpc}^{-2}$ and ranges from $1.7 \times 10^{16}\ h\ M_\odot\ {\rm Mpc}^{-2}$ to $3.9 \times 10^{15}\ h\ M_\odot\ {\rm Mpc}^{-2}$ over this range of source redshifts. If the FBG $N(z)$ shape follows the no-evolution model (as used in Glazebrook et al. 1995) then $\beta \simeq 0.22$ and $\Sigma_{\rm crit} = 8.8 \times 10^{14}\ M_\odot\ h\ {\rm Mpc}^{-2}$.

In Figure 8 we plot the cluster radial mass profile for three different values of β corresponding to the faint lensed galaxies lying on sheets at $z_s = 1$, 1.5, and 3. Also shown for comparison are isothermal sphere mass profiles with velocity dispersions 2200, 1450, and 1100 km s^{-1}. A con-

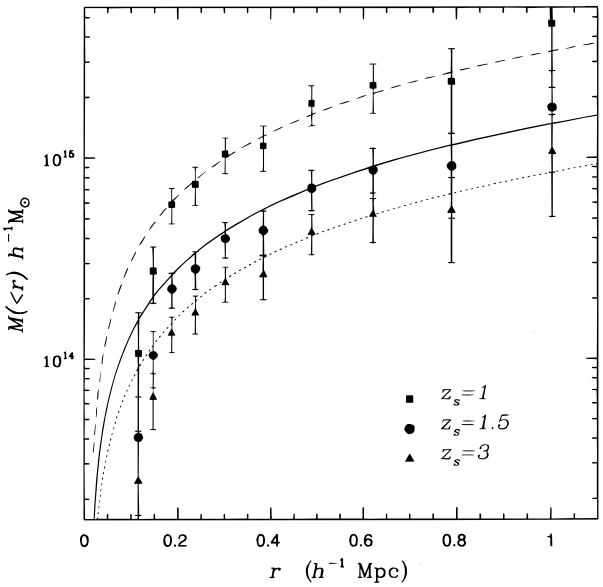

FIG. 8.—Plot of the radial mass profile $[M(<r) > \pi r^2\kappa(r)\Sigma_{\rm crit} = c^2r^2\kappa/(4GD_l\beta)]$ of MS 1054−03 using the κ (or ζ) values from the $I > 21.5$ sample for three different values of β assuming the faint lensed galaxies lie on sheets at $z_s = 1$, $z_s = 1.5$, and $z_s = 3$. The error bars reflect only the errors in κ and not the uncertainty in $\Sigma_{\rm crit}$. The dashed, solid, and dotted lines are mass profiles for isothermal spheres with $\sigma = 2200$, 1450, and 1100 km s^{-1}, respectively.

servative lower bound on the cluster mass is obtained if we assume that the faint/blue galaxies lie at $z_s = 3$, and we then find $M(<0.25) = (2.7 \pm 0.6) \times 10^{14}\ h^{-1}\ M_\odot$ and $M(<0.5) = (5.9 \pm 1.3) \times 10^{14}\ h^{-1}\ M_\odot$. For the no-evolution $N(z)$, $M(<0.5) = (1.39 \pm 0.29) \times 10^{15}\ h^{-1}\ M_\odot$.

We can combine these projected mass estimates with the projected light estimates of § 3 to obtain the cluster mass-to-light ratio. Since the mass estimates really measure the mean surface density in the aperture relative to that in the surrounding annulus we reduce the luminosity estimates by the expected mean surface brightness (this is a small correction; roughly 5% and 15% for the smaller and larger apertures, respectively). If we place the faint/blue galaxies at $z_s = 3$ then we obtain $M/L_V \simeq 250\ h$ for the small aperture and $M/L_V \simeq 350\ h$ for the larger (with $\simeq$21% statistical uncertainty). If instead they lie at $z_s = 1.5$, then the mass increases by roughly 70% and the mass-to-light ratio (for the 0.5 h^{-1}Mpc aperture) rises to $M/L_V \simeq 580$. For the no-evolution $N(z)$ we find $M/L_V = (790 \pm 170)\ h$ and for $z_s < 1$ we would require $M/L_V > 1600\ h$. Note that those values are the required mass-to-light ratios at $z = 0.83$. Passive-evolution models (Bruzual & Charlot 1993) predict a fading for the cluster galaxies of $\sim$1 mag or more since $z = 0.83$, making it even harder to reconcile our results with, for example, the no-evolution form for the FBG $N(z)$.

Finally, the net shear (which is sensitive to structures outside the beam) is $\gamma = \{0.019, -0.016\} \pm 0.012$, which is essentially a null detection, but at a precision level which is already at about the level of the expected signal from large-scale structure, so the prospects for constraining the large-scale mass power spectrum $P(k)$ with large angle surveys is excellent.

5. DISCUSSION

These results have implications for both the properties of high-z clusters (and therefore for cosmogonical theory) and for the $N(z)$ of the FBGs.

Regarding the cluster properties, we have found that the mass-to-light ratio is greater than 350 h, with the lower limit corresponding to having all the faint lensed galaxies at $z = 3$. This must be an underestimate as some of the galaxies surely lie at lower redshifts. For a more plausible mean redshift of, say $z_s = 1.5$, we obtain $M/L \simeq 580\ h$ (although a somewhat lower value for the central mass-to-light ratio), and for the no-evolution model $M/L \simeq 800\ h$. This is quite large compared to values normally obtained from the X-ray or virial analysis but is quite consistent with values measured by weak lensing for other lower redshift clusters (Fahlman et al. 1994; Smail, Ellis, & Fitchett 1995; Tyson & Fischer 1995; Squires et al. 1995).

The high M/L coupled with the high luminosity of the cluster makes it very massive indeed—it has the same projected surface mass density as a Navarro model (Navarro, Frenk, & White 1995) with rotation velocity v_{200} in the range 2400–2800 km s^{-1} or as an isothermal sphere with line-of-sight velocity dispersion of 1100–2200 km s^{-1} (see Fig. 8). The existence of large clusters like this at high redshift is problematic for hierarchical cosmological models like CDM with $\Omega = 1$. While this problem has been recognized for some time (Evrard 1989; Peebles et al. 1989; Gunn 1990), it has not been taken too seriously because of the lack of conclusive evidence that any of the few known high-z clusters were truly massive. We now have firm evidence for at least one such system. Using the Press-Schechter approximation, the predicted comoving number density of $10^{15}\ h^{-1}$ clusters at $z \sim 0.8$ in a standard CDM model ($\sigma_8 = 1.1$) is at least an order of magnitude lower than the number density at $z = 0$ (Vianna & Liddle 1995). But the existence of only one $10^{15}\ h^{-1}\ M_\odot$ cluster at $z \sim 0.8$ in the EMSS survey volume corresponds to a comoving number density of order $n \sim 5 \times 10^{-8}\ h^3$ Mpc^{-3} (Luppino & Gioia 1995), comparable to the "local" density $n\ (M > 10^{15}\ h^{-1}\ M_\odot) \sim 10^{-7}\ h^3$ Mpc^{-3} (White, Efstathiou, & Frenk 1993). In mixed dark matter models, the predicted abundance of massive clusters drops even more rapidly with redshift than in standard CDM.

The question of the $N(z)$ for the FBG population has been a matter of debate for some time. While some of the faint field galaxy population consists of low-redshift ($z < 0.5$) dwarfs, there remains the possibility that large, star-forming galaxies at $z > 1$ make up a significant fraction of the FBG excess counts, especially at faint magnitudes (Cowie, Hu, & Songaila 1995). There have been hints of this high-redshift component to the FBGs from lensing observations of lower redshift ($z < 0.5$) clusters (Fort et al. 1992; Kneib et al. 1994), and Smail & Dickinson (1995) have reported the detection of weak shear by a putative cluster surrounding the radio galaxy 3C 324 at $z = 1.2$. Furthermore, there is some weak lensing evidence for a $z \sim 1.5$ mass concentration coincident with a group of very faint galaxies that may be partly responsible for the lensing of Q2345+007 (Mellier et al. 1994; Fischer et al. 1994). On the other hand, as mentioned earlier, the failure of Smail et al. to detect lensing in Cl 1603+43 might lead one to the opposite conclusion. Our observation shows unequivocally that the lensed, faint background galaxies are predominantly blue, and that the majority of these in the range $23.5 < I < 25.5$ lie at redshifts of order unity or greater. Unfortunately we cannot be more precise without some independent estimate of the mass of the cluster. What we *can* say, however, is that either extreme case is very interesting. On one hand, if the cluster has a mass-to-light ratio at the lower limit of $\sim$350 h, then nearly all of the FBGs

must lie at very high redshift. On the other hand, to accommodate a more reasonable $N(z)$, such as a "no-evolution" model, requires a mass-to-light ratio of $\sim 800\ h$ and the cluster would then be exceptionally massive and should have an enormous velocity dispersion and X-ray temperature (at least in so far as the cluster is approximately spherical and relaxed).

It is clear, however, that detailed information on the FBG $N(z)$ is quite within reach. What is needed is a sample of five or 10 massive clusters at similar redshift to MS 1054−03, along with a reasonably complete spectroscopic sample to say $I = 23$. Clusters can be selected by high X-ray luminosity (as they are discovered in ongoing surveys), or better yet, by high X-ray temperature when such temperatures become available (e.g., see Donahue 1996). Although, as we have seen, it is difficult to detect the lensing in the brighter galaxies, with a number of lenses the statistics will improve and we should be able to determine the relative distances for the faint galaxies relative to the brighter ones, and then use the spectroscopic redshifts to tie down the overall scale. Ongoing spectroscopic surveys with the largest telescopes are now beginning to obtain spectra at the magnitude limits required here. Using the Keck Telescope, Cowie et al. (1996) have taken spectra of a sample of several hundred galaxies nearly complete to $I = 23$ ($K = 20$, $B = 24.5$). Interestingly, when they split their sample by color (at $B-I = 1.6$), they find that the blue galaxies divide into distinctly separate low-redshift ($z \sim 0.25$) and high-redshift ($z > 0.8$) populations, with the bulk of the faintest blue galaxies located at high redshift (see Figs. 18 and 20 in Cowie et al. 1996). Combining these observations with weak lensing, it should be possible to constrain the redshifts of galaxies that are several magnitudes fainter than will be accessible to spectroscopy even with 8–10 m telescopes in the foreseeable future.

It is a pleasure to thank Lev Kofman, Isabella Gioia, Ken Chambers, Doug Clowe, Megan Donahue, Mark Metzger, Karl Glazebrook, Neal Trentham, and Len Cowie for stimulation, help, and advice. We also thank the referee for useful comments that helped us clarify several areas of this paper.

APPENDIX

CALIBRATION OF THE SHEAR-POLARIZATION RELATION

It is physically plausible, and can be rigorously shown, that a gravitational shear γ will induce a change in the measured polarization,

$$e \rightarrow e' = e + \gamma P_\gamma \,, \tag{A1}$$

so that the expectation value for the polarization is proportional to γ. In order to make a quantiative estimate of the shear, and to measure differences in the shear between different subsamples as we do here, we need to determine the constant of proportionality P_γ, which we term the "shear polarizability."

There are several ways this can be done. One can empirically calibrate ground-based images by shearing deep *HST* images, convolving with the ground-based psf and measuring the polarization for a sample of objects that are chosen to be representative of those seen from the ground. The precision of this approach is limited by the relatively small number of galaxies at the relevant fluxes in the small number of deep *HST* fields now available. Another minor worry is that the *HST* filters differ somewhat from those used in the ground-based observations, so there may be slight systematic differences between the ground- and space-based color-selected samples. An alternative method (Wilson et al. 1996) is to use iteratively deconvolved ground-based images. An advantage of this technique is that it could in principle be applied to "self-calibrate" *HST* images themselves, which would otherwise require modeling of the faint galaxies.

Here we have used another calibration method. It is similar to that of Wilson et al. in that it does not require auxiliary observations with a higher resolution instrument, but differs in that no deconvolution is needed, and the polarizability we obtain is a simple linear function of the observed galaxy surface brightness:

$$P_\gamma = P_{\rm sh} - P^*_{\rm sh}/P_{\rm sm}/P^*_{\rm sm} \,, \tag{A2}$$

where $P_{\rm sh}$, $P_{\rm sm}$ are as defined in KSB, and the superscript $*$ denotes the values for a stellar object. A detailed justification and testing of equation (A2) is given elsewhere (Kaiser 1996). Here we sketch how this result is derived.

Consider a galaxy with intrinsic surface brightness f. If we observe this with an instrument with a perfectly circular psf g we obtain the "ideal observed image"

$$f_o = g \otimes f \tag{A3}$$

from which we can calculate the ideal polarization statistic $e = e(f_o)$. Now KSB have shown that if one introduces a small anisotropy of the psf by convolving g with a small but highly anisotropic kernel q (which might give a good model of guiding errors, etc.),

$$f_o \rightarrow f'_o = g' \otimes f = q \otimes g \otimes f \,, \tag{A4}$$

then the response of the polarization is

$$e \rightarrow e' = e + P_{\rm sm} p \,, \tag{A5}$$

where p is a measure of the psf anisotropy and the "smear polarizability" $P_{\rm sm}$ can be measured from the actual image f'_o: $P_{\rm sm} = P_{\rm sm}(f'_o)$. We have used this result above to correct for psf anisotropy. KSB also showed that if one shears the ideal image,

$$f_o \rightarrow f'_o = S_\gamma(g \otimes f) \,, \tag{A6}$$

where the shear operator S_γ is defined such that

$$(S_\gamma h)_r = h[(\delta_{ij} + \psi_{ij})r_j] , \tag{A7}$$

with

$$\psi_{ij} = \begin{bmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & -\gamma_1 \end{bmatrix} , \tag{A8}$$

then the polarization response is

$$e \rightarrow e' = e + P_{\rm sh}\gamma \tag{A9}$$

now to linear order in γ, and where again, at linear order, we can measure $P_{\rm sh}$ from the actual image $P_{\rm sh} = P_{\rm sh}(f'_o)$. Combining these, if we convolve f with a (possibly weakly anisotropic) kernel g' and then apply a weak shear,

$$f' \rightarrow S_\gamma(g' \otimes f) , \tag{A10}$$

then the response to the combined operation is

$$e' \rightarrow e_s + P_{\rm sh}\gamma + P_{\rm sm}p(g') \tag{A11}$$

now to linear order in both γ and p. Unfortunately, $P_{\rm sh}$ cannot be used as it stands to calibrate the $\gamma - e$ relation since it describes the response of the polarization to a shear applied *after* seeing (rather than before as in real lensing) and will tend to overestimate the response for small objects. However, it is easy to show that the real process of a shear followed by a convolution with a circular seeing kernel g is precisely equivalent to a convolution with a slightly anisotropic (anti)sheared kernel $g' = S_\gamma^{-1}g$ followed by a shear:

$$f_o \rightarrow f'_o = g \otimes (S_\gamma f) = S_\gamma[(S_\gamma^{-1}g) \otimes f] . \tag{A12}$$

If we now make the assumption that the effect on the polarization of a psf anisotropy caused by shearing is essentially identical to that induced by smearing the ideal psf with an appropriate kernel q (and for seeing caused by turbulence this is an excellent approximation), then

$$e \rightarrow e' = e + P_{\rm sh}\gamma + P_{\rm sm}p(S_\gamma^{-1}g) . \tag{A13}$$

Finally, we can express $p(S_\gamma^{-1}g)$ in terms of $P_{\rm sh}$ and $P_{\rm sm}$ measured for stellar objects since for an intrinsically pointlike object $e = e' = 0$ (as shear has no effect on a point source), so the last two terms in equation (A13) must cancel, yielding

$$p(S_\gamma^{-1}g) = -\gamma P^*_{\rm sh}/P^*_{\rm sm} . \tag{A14}$$

Therefore, for a nonstellar object,

$$e \rightarrow e' = e + \gamma(P_{\rm sh} - P_{\rm sm}P^*_{\rm sh}/P^*_{\rm sm}) , \tag{A15}$$

and so $P_\gamma = d(e' - e)/d\gamma$ is given by equation (A2) as advertised.

The "preseeing shear polarizability" P_γ defined here (to distinguish it from the "postseeing shear polarizability" $P_{\rm sh}$) has very reasonable asymptotic behavior: for very large objects $P_{\rm sm}$ (which scales as the inverse area of the object) is much less than $P_{\rm sh}$ (which depends only on the object shape), and $P_\gamma \rightarrow P_{\rm sh}$. For small objects, the extra negative term tends to decrease the response, and for stellar objects $P_\gamma \rightarrow 0$ as expected.

Finally, a few words are in order to describe how we actually compute and apply P_γ. The ratio $P^*_{\rm sh}/P^*_{\rm sm}$ in equation (A2) is actually calculated as the ratio of the average shear/smear polarizabilities for the stars (although one could equally well use a single moderately bright star). One could in principle calibrate the polarization for each galaxy individually, to obtain a set of shear estimates $\hat{\gamma} = e/P_\gamma$. However, this tends to be very noisy for small objects. Instead, what we do is use as our estimate of the shear $\hat{\gamma} = e/\langle P_\gamma \rangle$, where the average of the polarizability is taken over all the galaxies in the relevant subsample.

REFERENCES

Bahcall, N. 1981, ApJ, 247, 787
Bonnet, H., Mellier, Y., & Fort, B. 1994, ApJ, 427, L83
Bruzual, G., & Charlot, A. 1993, ApJ, 405, 538
Castander, F., Ellis, R., Frenk, C., Dressler, A., & Gunn, J. 1994, ApJ, 424, L79
Couch, W., Ellis, R., Malin, D., & MacLaren, I. 1991, MNRAS, 249, 606
Cowie, L., Hu, E., & Songaila, A. 1995, Nature, 377, 603
Cowie, L., Songaila, A., Hu, E., & Cohen, J. 1996, AJ, submitted
Donahue, M. 1996, preprint
Evrard, A. 1989, ApJ, 341, L71
Fahlman, G., Kaiser, N., Squires, G., & Woods, D. 1994, ApJ, 437, 56
Fischer, P., Tyson, J.A., Bernstein, G., & Guhathakurta, P. 1994, ApJ, 431, L71
Gioia, I., & Luppino, G. A. 1994, ApJS, 94, 583
Gioia, I. M., Maccacaro, T., Schild, R.E., Wolter, A., Stocke, J.T., Morris, S. L., Henry, J. P. 1990, ApJS, 72, 567
Glazebrook, K., Ellis, R., Colless, M., Broadhurst, T., Allington-Smith, J., & Tanvir, N. 1995, MNRAS, 273, 157
Guhathakurta, P., Tyson, J. A., & Majewski, S. 1990, ApJ, 357, L9
Gunn, J., Hoessel, J., & Oke, J. B. 1986, ApJ, 306, 30
Gunn, J. 1990, in Clusters of Galaxies, ed. W. Oergerle et al. (Cambridge: Cambridge Univ. Press), 341
Henry, J. P., Gioia, I. M., Maccacaro, T., Morris, S. L., Stocke, J. T., & Wolter, A. 1992, ApJ, 386, 408
Kaiser, N. 1995, ApJ, 439, L1
———. 1996, in preparation
Kaiser, N., Broadhurst, T., Szalay, A., & Moller, P. 1996, in preparation
Kaiser, N., & Squires, G. 1993, ApJ, 404, 441 (KS93)
Kaiser, N., Squires, G., & Broadhurst, T. 1995, ApJ, 449, 460 (KSB)
Kneib, J.-P., Mathex, G., Fort, B., Mellier, Y., Soucail, G., & Longaretti, P.-Y. 1994, A&A, 286, 701
Landolt, A. 1992, AJ, 104, 340
Luppino, G., & Gioia, I. 1995, ApJ, 445, L77
Mellier, Y., Dantel-Fort, M., Fort, B., & Bonnet, H. 1994, A&A, 289, L15
Navarro, J., Frenk, C., & White, S. 1995, MNRAS, 275, 720
Peebles, J., Daly, R., & Juszkiewicz, R. 1989, ApJ, 347, 563
Schneider, P. 1995, A&A, 302, 639
Smail, I., Ellis, R., Fitchett, M., & Edge, A. 1994, MNRAS, 270, 245
Smail, I., & Dickinson, M. 1995, ApJ, 455, L99
Smail, I., Ellis, R., & Fitchett, M. 1995, MNRAS, 273, 277
Squires, G., & Kaiser, N. 1996, ApJ, in press (SK96)
Squires, G., Kaiser, N., Fahlman, G., Woods, D., Babul, A., Neumann, D., & Bohringer, H. 1995, preprint
Tyson, J. A., & Fischer, P. 1995, ApJ, 446, L55
Tyson, J. A., Valdes, F., & Wenk, R. 1990, ApJ, 349, L1
Vianna, P., & Liddle, A. 1995, preprint
White, S., Efstathiou, G., & Frenk, C. 1993, MNRAS, 262, 1023
Wilson, G., Cole, S., & Frenk, C. S. 1996, MNRAS, 280, 199