First Identification of a CMB Lensing Signal Produced by 1.5 Million Galaxies at $z \sim 4$: Constraints on Matter Density Fluctuations at High Redshift

Hironao Miyatake,1,2,3,4 Yuichi Harikane,5,6 Masami Ouchi,7,5,4 Yoshiaki Ono,5 Nanaka Yamamoto,3 Atsushi J. Nishizawa,8,2,3 Neta Bahcall,9 Satoshi Miyazaki,7 and Andrés A. Plazas Malagón9

1Kobayashi-Maskawa Institute for the Origin of Particles and the Universe (KMI), Nagoya University, Nagoya 464-8602, Japan
2Institute for Advanced Research, Nagoya University, Nagoya 464-8601, Japan
3Division of Physics and Astrophysical Science, Graduate School of Science, Nagoya University, Furo-cho, Chikusa, Nagoya 464-8602, Japan
4Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU, WPI), UTIAS, The University of Tokyo, 5-1-5 Kashiwa-no-ha, Kashiwa, Chiba 277-8583, Japan
5Institute for Cosmic Ray Research, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8582, Japan
6Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, United Kingdom
7National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan
8DX Center, Gifu Shotoku Gakuen University, Takakuwa-Nishi, Yanaizuchō, Gifu, 501-6194, Japan
9Department of Astrophysical Sciences, Peyton Hall, Princeton University, Princeton, New Jersey 08540, USA

(Received 31 March 2021; revised 21 March 2022; accepted 9 May 2022; published 1 August 2022)

We report the first detection of the dark matter distribution around Lyman break galaxies (LBGs) at high redshift through the cosmic microwave background (CMB) lensing measurements with the public Planck PR3 $\kappa$ map. The LBG sample consists of 1473 106 objects with the median redshift of $z \sim 4$ that are identified in a total area of 305 deg$^2$ observed by the Hyper Suprime-Cam Strategic Survey Program survey. After careful investigations of systematic uncertainties, such as contamination from foreground galaxies and cosmic infrared background, we obtain the significant detection of the CMB lensing signal at 51e that is dominated by 2-halo term signals of the LBGs. Fitting a simple model consisting of the Navarro-Frenk-White profile and the linear-bias model, we obtain the typical halo mass of $M_h = 2.9^{+9.5}_{-2.5} \times 10^{11}$ h$^{-1}$ M$_\odot$. Combining the CMB lensing and galaxy-galaxy clustering signals on the large scales, we demonstrate the first cosmological analysis at $z \sim 4$ that constrains $(\Omega_m, \sigma_8)$. We find that our constraint on $\sigma_8$ is roughly consistent with the Planck cosmology, while this $\sigma_8$ constraint is lower than the Planck cosmology over the 1$\sigma$ level. This study opens up a new window for constraining cosmological parameters at high redshift by the combination of CMB and high-$z$ galaxies, as well as studying the interplay between galaxy evolution and large-scale structure at such high redshift, by upcoming CMB and optical and near-infrared imaging surveys.

DOI: 10.1103/PhysRevLett.129.061301

Introduction.—Understanding the interplay between galaxy evolution and large-scale structure is key to understanding cosmic evolution. Since galaxies are formed in dark matter halos through gas cooling, such an interplay can be studied by measuring the connection between dark matter halos and galaxies (see reviews by Refs. [1,2]). Galaxy-galaxy clustering, the autocorrelation of galaxy positions, is one of the statistical probes that has been widely used to investigate the galaxy-halo connection (e.g., Refs. [3–7]). Galaxy-galaxy lensing, the cross-correlation between galaxy positions and weak lensing shear of background galaxies, is rapidly emerging as another powerful probe because it enables the direct measurement of dark matter distribution around galaxies (e.g., Refs. [8–11]).

The combined measurement of galaxy-galaxy clustering and galaxy-galaxy lensing allows us to constrain cosmological parameters through growth of structure, and thus it is now the standard technique in cosmological galaxy surveys. Such studies can be divided into those using both small and large scales that enable us to investigate the galaxy-halo connection simultaneously (e.g., Refs. [12,13]), and those focusing on large scales to avoid possible systematics due to our incomplete understanding of galaxy physics (e.g., Refs. [13–16]). These measurements advanced our understanding of cosmic evolution of large-scale structure at $z < 1$ through the constraint on the amplitude of matter power spectrum $\sigma_8$, which is consistent with the Planck...
cosmology so far. As an independent test, it is of great importance to add such a measurement at higher redshift. Having structure growth measurements with a wide redshift may also allow us to address the $H_0$ tension between measurements of late time probes such as the type Ia supernova distance and early time probes such as cosmic microwave background (CMB) (e.g., Refs. [17,18]), through the sensitivity to early dark energy [19].

Recently, detailed study results of galaxy properties and galaxy-halo connection through clustering measurement of Lyman break galaxies (LBGs) at $z \sim 4$–7 are reported [7,20], where they used wide-field galaxy imaging survey data such as the Canada-France-Hawaii Telescope Legacy Survey and the Subaru Hyper Suprime-Cam (HSC) Subaru Strategic Program (SSP) Survey [21]. Although one of the next steps is measuring weak lensing on background galaxies around the LBGs, it is practically impossible due to the lack of data for shear measurements of galaxies behind LBGs. Instead, we can use weak lensing on the CMB anisotropies, which is sensitive to the large-scale structure at $z > 1$. In this Letter, using CMB lensing measurements in the Planck satellite data, we present the first detection of dark matter distribution around LBGs at $z \sim 4$. We also show cosmological constraints from clustering and lensing measurements.

Data: HSC GOLDRUSH catalog.—We use the second public data release (PDR2) product taken by the HSC SSP from March 2014 to January 2018 [22]. We use the wide layer data whose total full color full depth area is 305 deg$^2$. The HSC data are reduced by the HSC-SSP collaboration with HSCPIPE version 6.7 [23,24]. The typical $5\sigma$ limiting magnitudes in a 1.5-arcsecond diameter aperture are 26.4, 25.9, 25.8, 25.1, and 24.2 mag in $g$, $r$, $i$, $z$, and $y$ bands, respectively.

To construct a $z \sim 4$ LBG catalog of Great Optically Luminous Dropout Research Using Subaru HSC (GOLDRUSH) [25], we use the dropout selection method. Here, we briefly describe the selection method; full details of the catalog construction are presented in Harikane et al. [26]. We use the following color criteria to select $z \sim 4$ galaxies:

$$g - r > 1.0$$

(1)

$$r - i < 1.0$$

(2)

$$g - r > 1.5(r - i) + 0.8,$$

(3)

which are the same as the GOLDRUSH catalog construction with the first HSC data release [7,25,27]. In addition to these criteria, we apply a detection limit of a $> 5\sigma$ level in the $i$ band. We evaluate source detections with aperture magnitudes, and measure total fluxes and colors of sources with CONVOLVEFLUX magnitudes, which show good agreement with the CMODEL magnitude in the first public data release [28]. We mask out regions affected by bright object halos or unreliable background subtractions. Finally, we select a total of $\sim 2 \times 10^6$ LBGs. Ono et al. [25] used spectroscopically identified high-$z$ galaxies in the dropout sample to estimate the average redshift of the sample ($z = 3.8$ (median redshift of $z \sim 4$), which we adopt for our analysis. The surface number density and angular correlation functions of the LBGs agree well with previous studies. In our analysis, we use 1643863 LBGs, whose total $i$-band magnitudes are brighter than 25.5 mag, which are reduced to 1 473 106 LBGs after applying the Planck mask described below. Some foreground objects such as red galaxies at intermediate redshift can satisfy our selection criteria by photometric errors. A contamination rate is estimated to be $f_{\text{cont}} \sim 0.25$ based on HSC wide layer results obtained by [25], who derive a fraction of contaminating galaxies as a function of $i$-band magnitude by calculating the number of foreground galaxies that satisfy the color criteria.

Planck $\kappa$ map.—We construct the convergence map using lensing products in the third public data release from the Planck satellite mission (PR3) [29,30]. We use the minimum variance (MV) estimator and masks in the baseline lensing potential estimates. The MV estimators are provided as harmonic coefficients with the maximum multipole $L_{\max} = 4096$, while the mask is in real space. To construct the convergence map, we subtract the MV mean field from the MV convergence, and transform the harmonic coefficients to a real-space map applying a Gaussian filter with $\sigma = 2\degree$. We apply the mask to remove the galactic plane, point sources, and Sunyaev-Zel’dovich (SZ) clusters, which leaves a total unmasked sky fraction $f_{\text{sky}} = 0.671$. This results in the HEALPIX map with NSIDE $= 2048$ where the area of each healpixel is 2.95 arcsec$^2$. Note that the HSC-SSP survey does not include the galactic plane, and that we have a full overlap between our LBG sample and the Planck convergence map. After removing regions of point sources and SZ clusters masked in the Planck data, the total overlap area is 270 deg$^2$.

Measurement.—To measure the stacked convergence profile, we first create galaxy-centric angular bins for each galaxy by linearly dividing $0\degree < \theta < 20\degree$ into 14 bins, and identify healpixels that belong to each bin. Note that we validate the binning scheme as follows. First, we generate the 10 000 realizations of mock signals by fluctuating the convergence signal following the jackknife covariance described below. We then compute $\chi^2 = \sum_{ij} (\kappa_{\text{mock},i} - \kappa_j)(\text{Cov}^{-1})_{ij}(\kappa_{\text{mock},j} - \kappa_j)$ for each mock realization and confirmed that the $\chi^2$ distribution follows the predicted $\chi^2$ distribution with degree of freedom of 14.

We then collect the healpixels over all galaxies, and compute the average convergence in each bin. To subtract the residual mean field, we also perform the same measurement around random points, and subtract the random
signal from the observed convergence profile. The number of random points is 40 times larger than that of the LBGs.

We estimate the jackknife covariance dividing the PDR2 area into 262 subregions such that each subregion has a similar number of galaxies, and measure the mean-subtracted signal for each jackknife subregion. Note that the covariance is highly correlated due to the low-pass filter applied in the process of MV convergence estimation.

To test the contamination from cosmic infrared background, we stack Planck SZ-nulled temperature map (SMICA_NOSZ) around the LBGs. We do not find a gradient in the signal, and thus conclude that there is no significant cosmic infrared background contamination. In addition we confirm that there is no significant B-mode signal.

Results.—In Fig. 1, we show the measured convergence profile together with the mean field profile and the convergence profile before the mean field subtraction. We confirm that the observed amplitude of mean field signal is typical by measuring the convergence around random points using the publicly available simulated lens map in PR3. Defining the significance of the convergence profile after the mean field subtraction as \[ \frac{S/N}{\sigma} = \sum_{ij} k_i k_j \text{Cov}^{-1}_{ij} k_i, \] we obtain the significance level of \[ \frac{S/N}{\sigma} \approx 5.1 \] with all radial bins. Note that the convergence of each subfield of the PDR2 data [22] tends to be positive but we do not find a clear signal. This is because these subfields have a much smaller number of LBGs compared to our full sample.

The convergence profile is contaminated by foreground galaxies as described in the previous section. We thus model our signal as \[ \kappa^\text{obs}_{\text{tot}} = (1 - f_{\text{cont}}) \kappa^\text{obs}_{\text{LBG}} + f_{\text{cont}} \kappa^\text{obs}_{\text{cont}}. \] We use the halo models for the convergence profiles of real LBGs and contaminating galaxies with different halo masses and redshifts.

Let us first consider the convergence profile without the Planck beam. In the halo model approach, the convergence profile is composed of 1-halo and 2-halo terms: \[ \kappa(\theta) = k_1 h(\theta) + k_2 h(\theta), \] where \( k_1 h(\theta) \) and \( k_2 h(\theta) \) are the contribution from dark matter halos around galaxies and neighboring halos, respectively. We model the 1-halo term as \[ k_1 h(R) = \Sigma_{i} \frac{1}{(z_i, z_{\text{CMB}})} \frac{dA}{\theta} \int \frac{dk}{2\pi} \hat{u}_m(k; M_h, z_l) J_0(kR), \] where \( z_l \) (\( z_{\text{CMB}} \)) is the redshift of lens (CMB), \( \hat{u}_m \) is the mean matter density at present, \( \tilde{u}_m(k; z_l) \) is the Fourier-space Navarro-Frenk-White [31,32] profile truncated at the radius where the enclosed mass is 200 times the mean density \( R_200 \), \( J_0(x) \) is the zeroth-order Bessel function of the first kind. We adopt the analytical form of \( \tilde{u}_m(k; z_l) \) provided by Takada and Jain [33], and assume the concentration-mass relation derived by Duffy et al. [34]. The critical surface density is defined as \[ \Sigma_{\text{c}}(z_l, z_s) = c^2 (4\pi G)^{-1} d_A(z_l) d_A(z_l, z_s)^{-1} (1 + z_l)^{-2}, \] where \( c \) is the speed of light, \( G \) is the gravitational constant, \( d_A(z_l), d_A(z_l, z_s) \) are the angular diameter distances for the lens-source system. To model the 2-halo term, we adopt the linear bias model, i.e., \[ k_2 h(R) = \Sigma_{i} \frac{1}{(z_i, z_{\text{CMB}})} b(M_h, z_l) \hat{u}_m \int \frac{dk}{2\pi} P_m(k, z_l) J_0(kR), \] where \( b(M_h, z_l) \) is the linear bias parameter, and \( P_m(k, z_l) \) is the linear matter power spectrum at \( z_l \). We adopt the fitting function for the linear bias derived by Tinker et al. [35], and compute the matter power spectrum using CAMB [36]. Note that in our model the 1-halo and 2-halo terms are connected through halo mass. Finally, we convolve the Planck beam by applying the low pass filter \( L \leq 4096 \) and the Gaussian filter with \( \sigma = 2' \) to obtain a model for the observed signal \( \kappa^\text{obs}(\theta) \).

To estimate the halo mass of contaminating galaxies, we measure the weak lensing signal around the LBGs with the HSC first-year shear catalog [37,38]. This measurement enables us to pick up the lensing signal from contaminating galaxies, since most of the HSC source galaxies reside at \( z < 4 \) and thus are insensitive to the real LBGs at \( z = 3.8 \).

We first measure the excess surface mass density \( \Delta \Sigma(R) \), following the procedure described in Mandelbaum et al. [37]. Note that we include the random subtraction in the estimator, and we do not apply the boost factor correction since the boost factor is less than 1% [39]. Since we use the full probability density functions of photometric redshift (photo-z), the photo-z correction needed for the use of point estimates is not applied [40]. The covariance is estimated using the jackknife method using the same jackknife subsamples as the CMB lensing measurement. We fit the signal with a model based on halo occupation distribution with off-centering (5 parameters for halo occupation distribution and two parameters for off-centering), which is similar to the one described in More et al. [13], using the publicly available Markov Chain Mote Carlo (MCMC) sampler called EMCEE [41]. We translate the best-fit model in \( \Delta \Sigma \) into convergence \( \kappa \) to obtain the contamination signal.
We estimate the impact of the uncertainty associated with LBG redshifts by using a subsample of LBGs that have spectroscopic redshifts. Using the spectroscopic redshifts, we find the model varies by 4% and 17% for lensing and clustering, respectively. These deviations are much smaller than the statistical uncertainties of the measurements, which are typically 70% and 40% for lensing and clustering, respectively.

We constrain the cosmological parameters $\Omega_{m0}$ and $\sigma_8$ by simultaneously fitting the large-scale clustering and lensing signal, while the other cosmological parameters are fixed to the Planck cosmology [45]. We model the lensing signal in the same manner as the modeling procedures described in the previous section with the following modifications. For the convergence profile of LBGs, we use only the 2-halo term and vary $b$ as a fitting parameter. We also vary $\Omega_{m0}$ and $A_s$, on which the linear power spectrum depends. Both 1-halo and 2-halo terms in the lensing signal from contaminating galaxies are fixed. We model the observed clustering signal as $\omega_{\text{halo}}(\theta) = (1 - f_{\text{cont}})^2 \omega_{\text{LBG}}(\theta)$, where $\omega_{\text{LBG}}(\theta)$ is the clustering signal from LBGs. Note that we ignore the contamination signal for clustering signal because it is suppressed by $f_{\text{cont}}^2 \sim 0.06$. The projected clustering signal of LBGs is modeled as $\omega_{\text{LBG}}(\theta) = \int dz N^2(z) (dr/dz)^{-1} \Delta_k \int dk (k/2\pi)b^2 P_{\text{mm}}(k, z; \Omega_{m0}, A_s) J_0(r(z)\theta k)$, where $N(z)$ is the normalized redshift distribution of galaxies taken from [25].

We perform an MCMC analysis with flat priors, $\Omega_{m0} \in [0.01, 0.95]$, $10^{10} A_s \in [0.1, 5.0]$, and $b \in [0, 1.30]$. Figure 2 shows constraints on these parameters for the combined lensing and clustering results, together with constraints based on the lensing or clustering results alone. Since $\sigma_8$ and $b$ degenerate in a different manner for lensing and clustering at small $b$, the combination of lensing and clustering probes improves the constraint on $\sigma_8$, yielding $\sigma_8 = 0.46_{-0.25}^{+0.25}$ and $b = 6.5_{-4.2}^{+5.2}$ (68% credible level).

![FIG. 2. Cosmological constraints from CMB lensing and galaxy-galaxy clustering measurements at $z \sim 4$ (red contours). The blue (green) contours show constraints from the lensing (clustering) probe alone. The degeneracy between $\sigma_8$ and $b$ is reduced by the combination of these probes. These constraints are roughly consistent with the Planck cosmology [purple cross; [45]].](image)
Although it is known that there is a strong degeneracy between \( \Omega_{m0} \) and \( \sigma_8 \) in a combined analysis of lensing and clustering (e.g., Refs. [13,15,16]), the degeneracy is not observed in our analysis, and there is almost no constraint on \( \Omega_{m0} \). This is due to the fact that the sensitivity of matter power spectrum to \( \Omega_{m0} \) significantly decreases at high redshifts.

Figure 3 compares \( \sigma_8 \) constraints from our high-\( z \) study with those from low-\( z \) studies in the literature. Our constraint on \( f\sigma_8(z) \) is derived by converting our constraints on \( (\Omega_{m0}, \sigma_8) \) into \( f\sigma_8(z) \), where \( f = -d \ln D(z)/d(1 + z) \) is the logarithmic growth rate and \( \sigma_8(z) = \sigma_8 D(z)/D(0) \) is the linear matter fluctuation at redshift \( z \). Our constraints on both \( \sigma_8 \) and \( f\sigma_8(z) \) are roughly consistent with constraints computed from the Planck 2018 TT, TE, EE + lowE + lensing result [45], while our constraints are lower than the Planck cosmology at the 1.4\( \sigma \) level for both \( \sigma_8 \) and \( f\sigma_8(z) \). Motivated by the potential deviations from the Planck constraints on \( \Lambda \)CDM cosmology we repeat the same analysis with the assumption that the time dependent dark energy is characterized by the equation-of-state parameter \( w(a) = w_0 + (1 - a)w_a \). We find that again our constraint is consistent with the Planck cosmology.

The HSC-SSP survey will be completed by the end of 2021, which will cover \( \sim 1,400 \text{ deg}^2 \) of the sky and provide a GOLDRUSH sample three times as large as the one used in this Letter. The Advanced Atacama Cosmology Telescope Polarimeter (AdvACT) recently released the temperature and polarization maps much deeper than Planck, and will increase sensitivity in coming years [55,56]. Combining these datasets, we can significantly improve the signal-to-noise ratios of the lensing and clustering signals. In addition, the beam size of AdvACT is much smaller than Planck, which will improve systematics in CMB lensing measurements at small scales. This will enable us to perform better measurements of 1-halo terms. Moreover, in the near future, optical and near-infrared programs of the Rubin Observatory Legacy Survey of Space and Time (LSST) [57] and the Nancy Grace Roman Space Telescope [58] will provide much larger and cleaner LBG samples by their deeper and redder imaging. Spectroscopic observations planned with Roman will enable us to evaluate completeness and purity of the LBG samples. On the CMB side, Simons Observatory [59] and CMB-S4 [60], which will provide much deeper data compared to AdvACT while keeping the beam size, are scheduled to be deployed around the same time as the optical and near-infrared surveys start. The combination of these datasets will allow us to extend precision cosmology studies via the structure growth from high (\( z \sim 4 \)) to low redshift (\( z \sim 1 \)) [61]. The structure growth measurement at such a high redshift may enable us to explore a nonstandard cosmological model that can resolve the \( H_0 \) tension such as an early dark energy model [18]. In fact, Klypin et al. [19] showed that halo clustering at \( z \sim 4 \) (\( z \sim 6 \)) can be suppressed by \( \sim 20\% \) (\( \sim 30\% \)) by early dark energy. Thus measurements of the high-redshift large-scale structure can be one of the important probes for cosmology beyond the concordance model.

The Hyper Suprime-Cam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan, the Kavli Institute for the Physics and Mathematics of the Universe, the University of Tokyo, the High Energy Accelerator Research Organization, the Academia Sinica Institute for Astronomy and Astrophysics in Taiwan, and Princeton University. Funding was contributed by the FIRST program from the Japanese Cabinet.
of Education, Culture, Sports, Science and Technology, the Japan Society for the Promotion of Science, Japan Science and Technology Agency, the Toray Science Foundation, the National Astronomical Observatory of Japan, the Kavli Institute for the Physics and Mathematics of the Universe, the High Energy Accelerator Research Organization, the Academia Sinica Institute for Astronomy and Astrophysics, and Princeton University. This work makes use of software developed for the Large Synoptic Survey Telescope. We thank the LSST Project for making their code available as free software at http://dm.lsst.org. This work is based [in part] on data collected at the Subaru Telescope and retrieved from the HSC data archive system, which is operated by Subaru Telescope and Astronomy Data Center at the National Astronomical Observatory of Japan. Data analysis was in part carried out with the cooperation of Center for Computational Astrophysics at the National Astronomical Observatory of Japan. The Pan-STARRS1 surveys and public science archive have been made possible through contributions by the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg, and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, the Queen’s University Belfast, the Harvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central University of Taiwan, the Space Telescope Science Institute, the National Aeronautics and Space Administration under grant no. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation Grant No. AST-1238877, the University of Maryland, Eotvos Lorand University, the Los Alamos National Laboratory, and the Gordon and Betty Moore Foundation. We thank Mathew Madhavacheril, Shuichiro Yokoyama, and Kiyotomo Ichiki for useful discussions. This work was supported in part by World Premier International Research Center Initiative, the Ministry of Education, Culture, Sports, Science and Technology, Japan, and the Japan Society for the Promotion of Science KAKENHI Grants No. JP15H02064, JP19H05100, JP20H01932, and JP21H00070, and Nagoya University Research Fund.

*Corresponding author.
miyatake@kmi.nagoya-u.ac.jp


Throughout this Letter, we adopt the mode of the posterior distribution to infer the central value of the parameter, and the highest density interval of the posterior to infer the credible interval of the parameter. We use the marginalized posterior distribution when there are multiple fitting parameters.