Cosmological constraints from weak lensing scattering transform using HSC Y1 data

Sihao Cheng (程思浩)[1,2]★, Gabriela A. Marques[3,4], Daniela Grandón[5,6], Leander Thiele[7,8],
Masato Shirasaki[9,10], Brice Ménard[11], & Jia Liu[8]

[1]*Institute for Advanced Study, 1 Einstein Dr., Princeton, NJ 08540, USA*
[2]*Perimeter Institute for Theoretical Physics, 31 Caroline St N, Waterloo, ON N2L 2Y5, Canada*
[3]*Fermi National Accelerator Laboratory, P. O. Box 500, Batavia, IL 60510, USA*
[4]*Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA*
[5]*Mathematical Institute, Leiden University, Gorlaeus Gebouw, Einsteinweg 55, NL-2333 CC Leiden, The Netherlands*
[6]*Departamento de Física, FCFM, Universidad de Chile, Blanco Encalada 2008, Santiago, Chile*
[7]*Department of Physics, Princeton University, Princeton, NJ 08544, USA*
[8]*Center for Data-Driven Discovery, Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan*
[9]*National Astronomical Observatory of Japan (NAOJ), National Institutes of Natural Sciences, Osawa, Mitaka, Tokyo 181-8588, Japan*
[10]*The Institute of Statistical Mathematics, Tachikawa, Tokyo 190-8562, Japan*
[11]*Department of Physics and Astronomy, The Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218, USA*

26 April 2024

ABSTRACT

As weak lensing surveys go deeper, there is an increasing need for reliable characterization of non-Gaussian structures at small angular scales. Here we present the first cosmological constraints with weak lensing scattering transform, a statistical estimator that combines efficiency, robustness, and interpretability. With the Hyper Suprime-Cam survey (HSC) year 1 data, we obtain $\Omega_m = 0.29^{+0.04}_{-0.03}$, $S_8 \equiv \sigma_8(\Omega_m/0.3)^{0.5} = 0.83 \pm 0.02$, and intrinsic alignment strength $A_{IA} = 1.0 \pm 0.4$ through simulation-based forward modeling. Our constraints are consistent with those derived from *Planck*. The error bar of Ω_m is 2 times tighter than that obtained from the power spectrum when the same scale range is used. This constraining power is on par with that of convolutional neural networks, suggesting that further investment in spatial information extraction may not yield substantial benefits.

We also point out an internal tension of S_8 estimates linked to a redshift bin around $z \sim 1$ in the HSC data. We found that discarding that bin leads to a consistent decrease of S_8 from 0.83 to 0.79, for all statistical estimators. We argue that photometric redshift estimation is now the main limitation in the estimation of S_8 using HSC. This limitation is likely to affect other ground-based weak lensing surveys reaching redshifts greater than one. Alternative redshift estimation techniques, like clustering redshifts, may help alleviate this limitation.

Key words: methods: statistical – gravitational lensing: weak – cosmological parameters – large-scale structure of Universe

1 INTRODUCTION

Due to gravitational lensing, matter density fluctuations in our universe distort the images of background galaxies. On cosmological scales, this phenomenon is referred to as cosmic shear (for a review, see Kilbinger 2015; Mandelbaum 2018). This effect can be used to probe the geometry and the matter distribution of our Universe. At relatively large angular scales, the lensing field is approximately Gaussian and can be characterized by two-point statistics, such as the power spectrum or correlation function. However, as illustrated in Figure 1, ongoing weak lensing surveys can probe scales small enough to reveal rich non-Gaussian structures, requiring additional statistics to extract the full cosmological information in the field.

As opposed to Gaussian fields, non-Gaussianities possess much more degrees of freedom, and different statistics may exhibit sensitivity to diverse features and systematic effects in a map. The efficacy

of non-Gaussian statistics has been explored in the weak lensing context, including on the most recent datasets (Martinet et al. 2018; Fluri et al. 2022; Zürcher et al. 2022; Secco et al. 2022b; Liu et al. 2023; Thiele et al. 2023; Anbajagane et al. 2023; Heydenreich et al. 2022; Burger et al. 2023; Marques et al. 2024a). However, there has been no simple agreement on which summary statistic to use. Traditional moment based statistics are not robust to outliers in real data. Other simple statistics are usually too focused on particular features, and thus cannot extract the full cosmological information in a field. On the other extreme, inference with deep learning methods provide better constraining power but still lack interpretability.

Recently, a new summary statistic called the scattering transform has been proposed to extract cosmological information from weak lensing maps (Cheng et al. 2020). Similar statistics using related ideas, such as the phase harmonics (Allys et al. 2020) and scattering spectra (Cheng et al. 2023), have also been proposed and studied on mock data for cosmology. Originating from the computer vision community (Bruna & Mallat 2013), the scattering transform is
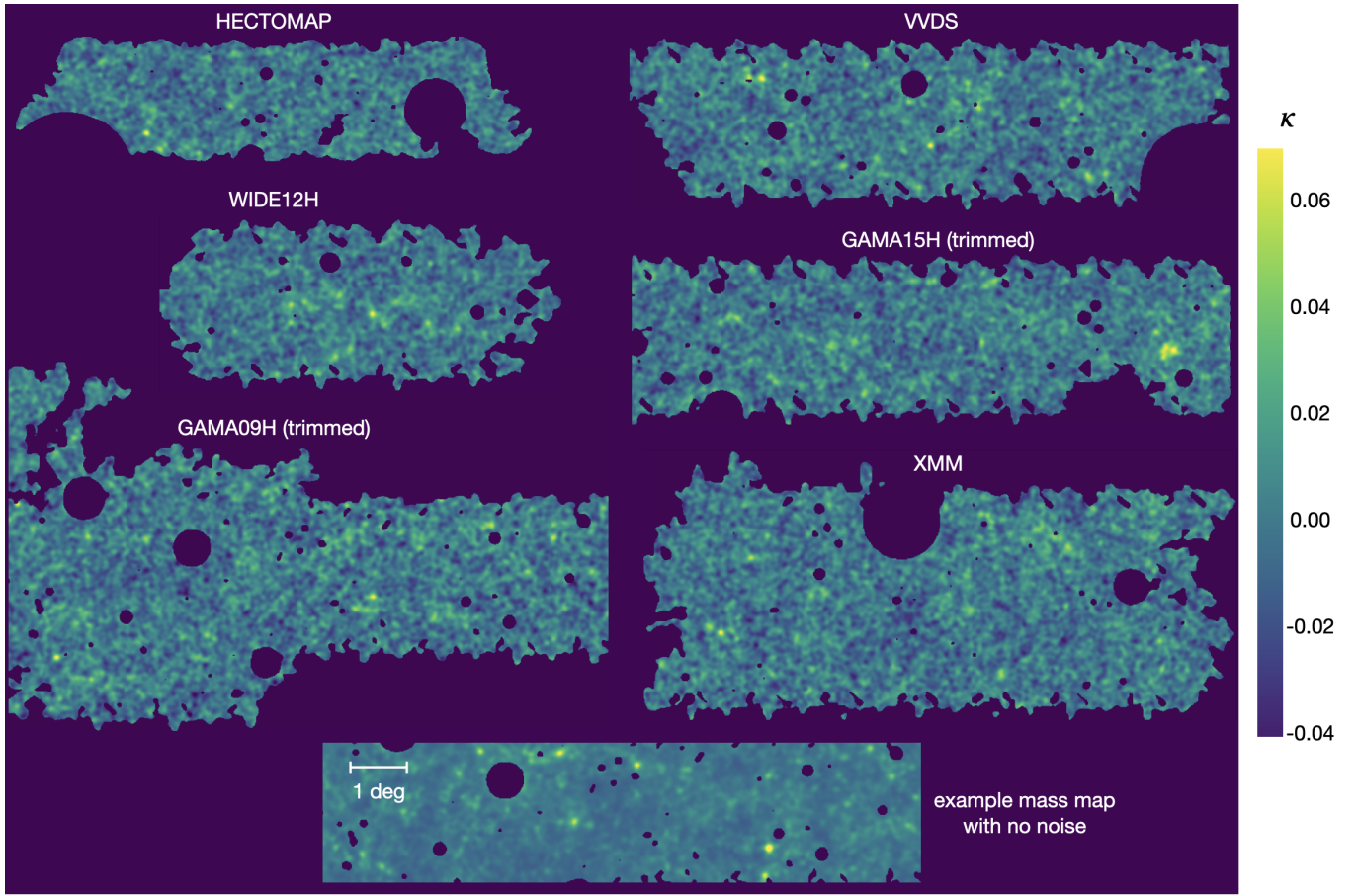
★ E-mail: scheng@ias.edu

Figure 1. The convergence maps made from HSC Y1 galaxy shear catalog (six panels), and a part of example mock map tailored to GAMA09H with no galaxy shape noise for reference. To visually demonstrate the cosmic shear signal, we only show maps using galaxies in the full photometric redshift range of 0.3–1.5 and thus with the highest signal-to-noise ratio. The maps are made using the procedure described in section 2.1 with 2 arcmin Gaussian smoothing. There are visible non-Gaussian structures, such as hierarchically clustered peaks above the noise level.

shown to exhibit high information content and robustness in mock cosmological (Cheng & Ménard 2021b; Gatti et al. 2023) and astrophysical (Allys et al. 2019) analysis. In this paper, we present the first application of this statistic to observational weak lensing data[1]. Our analysis uses the Subaru Hyper Suprime-Cam first-year (HSC Y1) data. We employ a forward modeling approach with a large set of N-body simulations and mock catalogs customized for the data, and derive cosmological constraints.

Weak lensing data is mainly sensitive to two cosmological parameters and provides roughly uncorrelated constraints on $S_8 \equiv \sigma_8 \sqrt{\Omega_m/0.3}$ and Ω_m. Here, Ω_m is the total matter density today, and σ_8 represents the linear matter fluctuation within a sphere of radius $8h^{-1}$Mpc. The scattering transform and in general non-Gaussian statistics are expected to significantly tighten the constraint of Ω_m by distinguishing fields with similar amplitude of lensing fluctuations but different extent into the non-linear regime. In addition, in light of the recent tension between S_8 values derived from the cosmic microwave background (Planck Collaboration

et al. 2020) and large-scale structures found in various studies (e.g., Hildebrandt et al. 2020; Asgari et al. 2021a; García-García et al. 2021; Abbott et al. 2022; Amon et al. 2022; Li et al. 2023; Dalal et al. 2023; Marques et al. 2024b), the exploration of non-Gaussian statistics provides a consistency check and may open a new avenue for understanding the underlying cause.

2 STATISTICAL ANALYSIS OF HSC Y1 DATA

We perform a statistical analysis of the first year data of the HSC survey[2] (Aihara et al. 2018), which is a wide-area imaging survey conducted using the 8.2m Subaru telescope in Hawaii in five broadbands, *grizy*. The HSC is the the deepest large-area weak lensing survey to date, with 5σ point-source depth of $i \sim 26$ and typically i-band seeing of 0.58". Hence, the HSC data is well suited for studying the non-Gaussian structures in the lensing field.

The first year data release of HSC, referred to as HSC S16A (Mandelbaum et al. 2018a), contains data taken from March 2014 to April 2016 and covers an area of 136.9 deg² area in six disjoint fields: GAMA09H, GAMA15H, HECTOMAP, VVDS, WIDE12H, and XMM.

---

[1] Application of the scattering transform has been explored to galaxy clustering with BOSS data (Valogiannis & Dvorkin 2022; Valogiannis et al. 2023) but not yet on weak lensing data.

[2] https://hsc.mtk.nao.ac.jp/ssp/

Due to the limited 10×10 deg^2 size of our cosmo-varied simulations, we trim off some edge regions in `GAMA09H` and `GAMA15H` that exceed a width of 10 degrees.

2.1 Weak lensing mass map

Our analysis is based on weak lensing convergence maps (also called mass maps) created using the galaxy shapes measured by Mandelbaum et al. (2018a) from the *i*-band coadded images with the re-Gaussianization PSF correction method (Hirata & Seljak 2003). Figure 1 shows the resulting mass maps, which clearly reveals non-Gaussian structures such as peaks and ridges. Below, we describe the procedures for creating the mass maps.

To perform a tomographic lensing analysis, we select source galaxies in 4 redshift bins, with edges [0.3, 0.6, 0.9, 1.2, 1.5]. Below, we will refer to those bins as bin 1 to 4, from low to high redshift. The redshift bin assignment is determined using the "best-fit" point estimate of redshift provided by the `Ephor-AB` photo-z code (Tanaka et al. 2018). The effective number density of galaxies in each photo-z bin is 5.14, 5.23, 3.99, 2.33 arcmin^{-2}, respectively. The total number of galaxies is about 8.5 million.

We begin by creating shear maps from the galaxy shape catalog. The survey area is discretized into pixels, each with a size of 0.88 arcminutes. Then, we calibrate and average the observed galaxy ellipticity e to obtain an unbiased estimator $\hat{\gamma}$ of the underlying cosmic shear γ for each pixel,

$$\hat{\gamma} = \frac{1}{1+\bar{m}_{\text{tot}}}\left(\frac{\bar{e}}{2\bar{\mathcal{R}}} - \bar{c}\right), \tag{1}$$

where $1+m_{\text{tot}}$ is the total multiplicative bias, c is the additive bias for each galaxy, and $\mathcal{R} \equiv 1-e_{\text{rms}}^2$ denotes the responsivity. This estimator originates from taking expectation of both sides of the shear distortion equation at small shear limit: $e = e^{\text{int}} + 2(1+m)\mathcal{R}\gamma + c + e^{\text{mea}}$, where e^{int} and e^{mea} are the intrinsic galaxy ellipticity and measurement error, both assumed to have zero expectation.

Following Hikage et al. (2019), we include three terms in the total multiplicative bias $m_{\text{tot}} = m + m_{\text{sel}} + m_{\mathcal{R}}$, where m is that obtained from image simulations (Mandelbaum et al. 2018b) and averaged within each of the individual fields and for each of the four redshift bins, $m_{\text{sel}} = [0.0086, 0.0099, 0.0091, 0.0091]$ is the multiplicative bias caused by galaxy size selection for each redshift bin, and $m_{\mathcal{R}} = [0.000, 0.000, 0.015, 0.030]$ is the correction for redshift-dependent responsivity (see section 5.3 of Mandelbaum et al. 2018b).

In eq. 1, the bar $\bar{\cdot}$ represents averages weighted by w:

$$\bar{\phi}(x, y) \equiv \frac{\sum_{\text{in pixel (x,y)}} w_i \phi_i}{\sum_{\text{in pixel (x,y)}} w_i}, \tag{2}$$

where $\phi(x, y)$ represent the value of any quantity in $\{e, \mathcal{R}, m_{\text{tot}}, c\}$ at the pixel (x,y), and the weights w_i are read directly from the HSC S16A catalog (Mandelbaum et al. 2018a), where it is defined as $w = (e_{\text{rms}}^2 + \sigma_e^2)^{-1}$, where σ_e^2 is the mean square of galaxy ellipticity in each component, estimated from all galaxies with similar color, brightness, and seeing, and σ_e^2 is the measurement error (Mandelbaum et al. 2018b). We further smooth our shear map with a Gaussian smoothing kernel of 2 arcmin in size to reduce the effect caused by the non-uniform weight map (galaxy counts per pixel). For Gaussian statistics, an uneven weight map can be analytically incorporated in the definition of the statistics. However, for non-Gaussian statistics except moment based ones, there is no general method to analytically incorporate the weight map, therefore a proper smoothing is usually needed. We note that compared to our two companion papers (Thiele et al. 2023; Marques et al. 2024a), the smoothing algorithm used in this paper incorporates the weight map in a way closer to the optimal weighting in power spectrum analysis and thus provides lower noise level. The algorithm of smoothing and more discussions are presented in appendix A. After smoothing, pixels with galaxy density lower than half of the mean density are masked and set to zero. For visualization purpose, we in-paint the masked regions (Pires et al. 2009) when generating Figure 1, but keep the masked regions as zero for statistical analysis. The shear maps are then converted into the mass maps using their arithmetic relation in Fourier space (Kaiser & Squires 1993) and implemented assuming periodic boundary conditions for each field of view.

2.2 Scattering transform and power spectra

We consider two types of summary statistics of the convergence maps: the power spectrum extracting Gaussian information, and scattering transform coefficients providing substantial non-Gaussian information.

The scattering transform was recently introduced to cosmology and has desirable properties when performing statistical analyses: it compresses the information in a robust way (Cheng et al. 2020). It achieves the constraining power of a deep convolutional neural network without the need to train millions of neurons. The number of coefficients used is only comparable to that of the power spectrum, which is smaller than most other non-Gaussian statistics. In practice, such an internal compression of information significantly reduces the estimation noise of the covariance matrix and likelihood emulator. Compared to high-order moment based statistics, the scattering coefficients do not amplify outliers, making the estimator stable under cosmic variance and robust to outliers in real observations.

The scattering transform also serves as a bridge between traditional moment-based statistics and convolutional neural networks. In the context of weak lensing maps, it can measure the scale-dependent sparsity of structures in the field, which reflects how non-linearly the fluctuations have evolved and therefore significantly improve the constraint of Ω_m. For a more intuitive understanding, we refer the reader to our pedagogical review (Cheng & Ménard 2021a).

From a lensing convergence field κ, the scattering transform first separates scales by wavelet convolutions, and then applies a point-wise modulus as non-linearity:

$$I_1^{j,\theta}(x, y) = |\kappa * \psi^{j,\theta}|, \tag{3}$$

where '∗' represents convolution, and wavelets ψ are band-pass filters labeled by their size j and orientation θ. As shown in Figure 2, we use wavelets that are one-sided in Fourier space and thus complex-valued in real space. The modulus therefore converts fluctuations into their local amplitude (i.e. envelope) so that the I_1 fields are local intensity maps of fluctuations in a frequency range. In particular we use directional Morlet wavelets, whose profiles in frequency space are close to off-centered Gaussians. The mathematical form is given in Appendix B.

The 1st-order scattering coefficients are the average of these I_1 fields over space and orientations

$$s_1(j) = \langle I_1 \rangle_{x,y,\theta}. \tag{4}$$

The s_1 coefficients are qualitatively similar to the power spectrum at the central frequency of the wavelet ψ^j – they just differ in the sense of being L_1-norm of the I_1 fields instead of the square of L_2-norms.

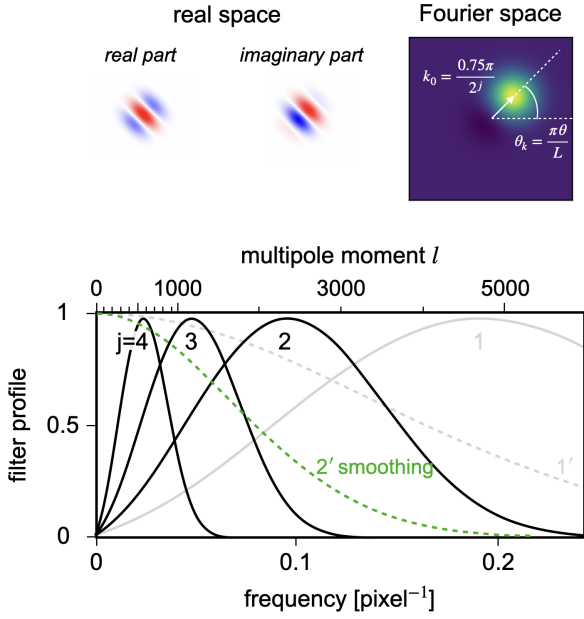To characterize more non-Gaussian features, each of the I_1 fields

Figure 2. *Upper*: The 2D profile of a Morlet wavelet used in the scattering transform. In real space it is a local wave packet, and in Fourier space it is a band pass window. Other wavelets in the same family can be created by dilation and rotation. *Lower*: The radial profile of a family of Morlet wavelets in Fourier space. The pixel size is 0.88 arcmin. In our main analysis, we use maps with 2 arcmin smoothing and wavelets with j=2–4.

are further analyzed following the same procedure,

$$I_2^{j_1,\theta_1,j_2,\theta_2}(x,y) = |I_1^{j_1,\theta_1} * \psi^{j_2,\theta_2}| = ||\kappa * \psi^{j_1,\theta_1}| * \psi^{j_2,\theta_2}| \quad (5)$$

$$s_2(j_1,j_2) = \langle I_2 \rangle_{x,y,\theta_1,\theta_2} \quad (6)$$

For the second-order scattering coefficients, as pointed out in (Cheng et al. 2020), most of the cosmological information lies in the orientation-average coefficients. Therefore, we average over both orientations θ_1 and θ_2. It has also been shown that only coefficients with $j_1 < j_2$ are informative. Intuitively, this is because the modulus converts fluctuations into their local amplitude (envelope), which tends to be smoother than the original fluctuations. The number of scattering coefficients is therefore $J + J(J-1)/2$, where J is the number of wavelet scales used.

Now we consider the power spectrum which carries only the Gaussian information. To cover exactly the same frequency range and make a fair comparison between the Gaussian and non-Gaussian information, we use the same wavelets ψ to weigh and bin the power spectrum, thereby obtaining a wavelet power spectrum,

$$C_l^{\text{wavelet}}(j) = \langle |\kappa * \psi^{j,\theta}|^2 \rangle_{x,y,\theta} = \langle I_1^2 \rangle_{x,y,\theta} \quad (7)$$

The wavelet power spectrum $C_l^{\text{wavelet}}(j)$ is the averaged power spectrum C_l within the frequency coverage of the wavelets[3]. For brevity, we shall use C_l to represent $C_l^{\text{wavelet}}(j)$ in figures and tables and specify the j-range in the captions. Compared to the normal binning of power spectrum, the wavelet binning guarantees that the scale coverage of the power spectrum is exactly the same as the scattering transform. We also compute the normally binned power spectrum,

---

[3] Strictly speaking, the averaging kernel is $\langle \int (\tilde{\psi}^{j,\theta}(k,\theta_k))^2 k \, d\theta_k \rangle_\theta$, while Figure 2 shows $\tilde{\psi}^{j,\theta}(k,\theta_k=0)$. More discussion can be found in the appendix C of Cheng & Ménard (2021a).
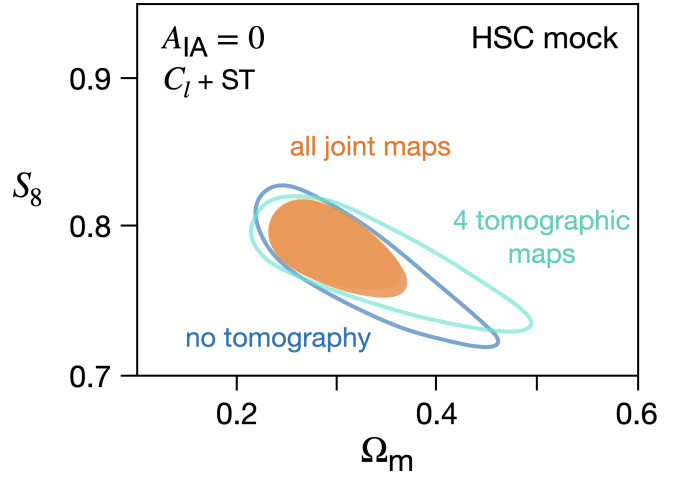
Figure 3. Significant improvement of constraining power by including joint non-Gaussian tomography, shown with mock inference. For simplicity we assume no intrinsic alignment. The orange region represents our default setting with 15 joint κ maps for scattering transform and 10 maps for the wavelet power spectra (see section 2.4 for more explanations) with j=2–4. It is the 68% contour of posterior inferred from the mean data vector at fiducial cosmology. The dark blue contour represents the constraint with one map using galaxies with all redshifts, and the cyan contour represents the constraint with four tomographic maps for galaxies from each individual redshift bin.

where 8 bins are set in the range $300 < l < 3500$ with equal spacing in logarithmic of l. They are computed directly from the κ map without further treatment on masks. As we follow a forward modeling approach and compute the statistics for observational and mock data in the same way, the mask should not bias the inference. When computing the scattering transform and power spectra, we adopt the flat sky approximation and assume periodic boundary condition. To reduce the influence of masks, when calculating the scattering transform s_1, s_2 and wavelet power spectrum $C_l(j)$, we exclude the masked regions in the last step of spatial average, i.e., $\langle \cdot \rangle_{x,y}$ becomes a weighted average with the mask as its weight.

2.3 Mock data and emulator

We follow a simulation-based inference approach to infer cosmological parameters from the observational data, as there is no analytical predictions for the scattering transform statistics so far. To build an emulator for the summary statistics, we use two sets of N-body simulations (Takahashi et al. 2017; Shirasaki et al. 2021) to generate multiple realizations of mock shear catalogs customized for our observational data (HSC Y1), and then make mass maps following the same pipeline as on real data.

Detailed description of the procedures to make the mock catalogs, the cosmological inference framework, and inclusion of systematics are presented in Appendix C, and details about the N-body simulations are presented in Appendix D. Below we only briefly summarize the information.

In our mock dataset, the N-body simulations assume the ΛCDM model. One set is used to sample cosmic variance at a fiducial cosmology (Takahashi et al. 2017), and the other is used to explore the cosmological dependence of Ω_m and S_8 (Shirasaki et al. 2021), which samples 100 cosmology in the range of $0.1 < \Omega_m < 0.7$ and $0.23 < S_8 < 1.1$. After the mock mass maps are made, we build a likelihood emulator of the summary statistics to perform cosmological

statistics	remarks	scale range	# of coefficients	$\Omega_m \times 1000$	$\sigma(\Omega_m) \times 1000$	$S_8 \times 1000$	A_{IA}	p-value
C_l + ST		j=2–4	120	**287^{+42}_{-30}**	**36**	**832 ± 18**	**0.96 ± 0.41**	**0.70**
ST		j=2–4	90	318^{+62}_{-41}	52	821^{+19}_{-22}	1.14 ± 0.40	0.46
C_l		j=2–4	30	300^{+83}_{-70}	77	820^{+18}_{-22}	1.27 ± 0.47	0.42
C_l		l=300–3500	80	357^{+126}_{-86}	106	814^{+22}_{-30}	1.21 ± 0.48	0.92
C_l		l=300–1900	60	290^{+152}_{-83}	118	814^{+21}_{-29}	1.00 ± 0.55	0.83
	no IA			252^{+25}_{-21}	23	831 ± 17	0	0.43
	1' smoothing			275^{+38}_{-27}	33	835 ± 16	0.76 ± 0.40	0.82
	TNG baryon			301^{+48}_{-33}	41	836 ± 19	0.95 ± 0.41	0.69
C_l + ST	no psf	j=2–4	120	299^{+47}_{-32}	39	834 ± 18	0.92 ± 0.41	0.69
	`mizuki` photo-z			294^{+48}_{-34}	41	825 ± 19	1.26 ± 0.41	0.59
	`frankenz` photo-z			262^{+70}_{-42}	56	829 ± 18	0.82 ± 0.33	0.57
	no bin 3			370^{+120}_{-79}	100	786^{+29}_{-36}	0.82 ± 0.36	0.53

Table 1. Summary of our cosmological constraints (marginal medians and 16 and 84 percentiles) using different summary statistics, and the p-values at the best fit of the posterior assuming Gaussian sampling distribution. The default setting (first line with boldface) uses 2 arcmin smoothing, `MLZ` photo-z and no baryonic effect.

inference. We fit the simulations with Gaussian sampling distribution whose mean vector changes with Ω_m and S_8, but covariance matrix is independent of cosmological parameters.

We also include the intrinsic alignment strength A_{IA} as a nuisance parameter to be inferred, and we marginalize over two systematic effects: multiplicative bias and PSF residual. For baryonic feedback and different photometric redshift, we only test their influence on different prescriptions but not marginalizing them. We do not include source clustering in our mock data as it is not expected to bias our statistics, which are even functions of the field values (e.g., Gatti et al. 2024).

2.4 Non-Gaussian tomography

Correlated Gaussian information between different tomographic maps can be exhausted by computing all the 4 auto- and 6 cross-power spectra. However, when considering non-Gaussian statistics, because many non-Gaussian statistics are defined on single maps, the analysis has usually been performed on individual tomographic maps, and there has not been an agreed way to extract correlated non-Gaussian information. In the context of peak count statistics, Martinet et al. (2021) proposed to include the cross information by simply computing non-Gaussian statistics on joint maps,

$$\kappa_{joint} = \sum_i N_i \kappa_i \qquad (8)$$

where N is the effective galaxy number density for a tomographic map and the summation is over subsets of $1, 2, 3, 4$. There are 11 such combinations plus the 4 individual maps. We point out that this is also similar to the way neural networks deal with multi-channel images, and it can be interesting to consider not only the summation, but more general linear combinations, and further look for the optimal choices. Nevertheless, in this study we just follow the idea of Martinet et al. (2021) and compute the scattering transform on the 15 maps. We find that including the cross-map information significantly improves the constraining power on Ω_m and A_{IA}. Figure 3 shows the mock inference for Ω_m and S_8 with no intrinsic alignment. Compared to the usual strategy of using only 4 tomographic maps or only 1 map with galaxies of all redshifts, a factor of two is earned by including the joint maps. In fact, when assuming $A_{IA} = 0$, it suffices to add only 3 joint maps, $\kappa_{1+2+3+4}$, κ_{2+3+4}, and κ_{3+4} to the 4 individual maps κ_1, κ_2, κ_3, and κ_4, which provides essentially the same constraining
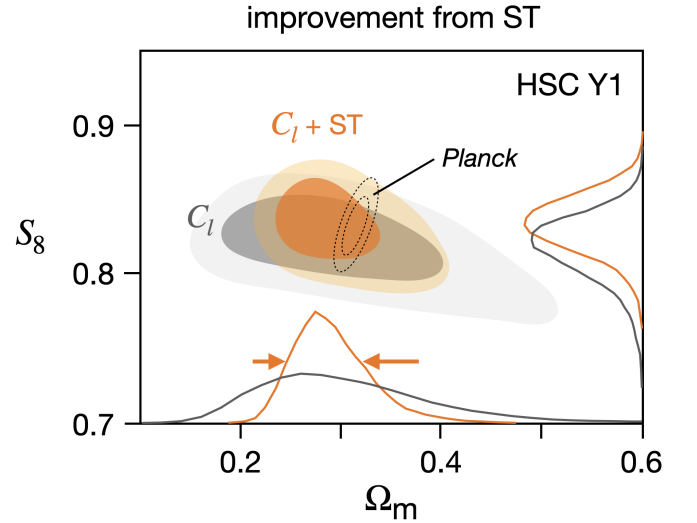

Figure 4. Posterior contours (68% and 95%) of Ω_m and $S_8 = \sigma_8 (\Omega_m/0.3)^{0.5}$ with HSC Year 1 data, using the tomographic scattering transform (ST) and wavelet power spectra covering the same frequency range (j=2–4), with intrinsic alignment parameter A_{IA} marginalized. The maps are smoothed by a 2' Gaussian kernel. The values of marginal constraints are listed in Table 1. The scattering transform tightens the constraints of Ω_m by a factor of 2.

power as using all 15 maps. However, when marginalizing over A_{IA}, including all the 15 maps improves the constraint of Ω_m by about 20%. A quantitative comparison using real HSC Y1 data is shown in Table 1.

3 RESULTS AND DISCUSSIONS

Figure 4 shows the cosmological constraints on HSC Y1 data. As the default setting, we use maps smoothed by a 2 arcmin Gaussian kernel and wavelets with $j = 2, 3, 4$ corresponding to central frequencies at $l = 600, 1200, 2400$ and together covering $l \approx 300$–3500 (see Figure 2). In total 90 scattering coefficients (45 s_1 and 45 s_2) are computed from 15 tomographic maps and joint maps. The wavelet power spectra is computed using the same wavelets for frequency

binning and on 10 tomographic and joint maps, which is enough to cover all Gaussian information.

The posteriors of Ω_m and S_8 are obtained with the intrinsic alignment parameter A_{IA} and other nuisance parameters marginalized. The constraints are also listed in Table 1. Our tight constraint obtained from the scattering transform and power spectra, $\Omega_m = 0.287^{+0.042}_{-0.030}$ and $S_8 \equiv \sigma_8(\Omega_m/0.3)^{0.5} = 0.832 \pm 0.018$, is consistent with the value inferred by Planck Collaboration et al. (2020) (TT,TE,EE+lowE+lensing) in both parameters. The small error bar of Ω_m and its consistency with *Planck* demonstrate the power and reliability of the scattering transform statistics.

In Table 1, we also show the constraints with several adjustments to the default setting to demonstrate the robustness of our result. In particular, we find that an Illustris-TNG baryonic feedback only shifts the S_8 inference up by 0.004 (0.2σ). Further study of the impact of baryons can be found in Grandón et al. (2024). Different photo-z codes or PSF-residual modeling do not shift S_8 significantly, either.

3.1 Non-Gaussian estimate improvements

By adding the scattering transform coefficients to the wavelet power spectra, the constraint on Ω_m is improved by a factor of 2[4]. Cheng et al. (2020) have shown that on a single-redshift convergence map, the constraining power of the scattering transform is on par with that of the convolutional neural networks reported in Ribli et al. (2019). Here we confirm the same constraining power on tomographic analysis of real data by comparing to Lu et al. (2023)'s convolutional neural net (CNN) results with the same data set.

Recently, new types of neural-net-based models have also been proposed for cosmological inference (Dai & Seljak 2022, 2024). In particular, the multiscale flow model combines the wavelet decomposition idea as used in the scattering transform with the normalizing flow idea to model the likelihood of the whole field, and has been reported to achieve better constraining power than CNN and the scattering transform on simulated weak lensing data (Dai & Seljak 2024). However, its performance in real data with source clustering, survey mask, and other systematics remains to be confirmed.

As for the constraining power of S_8, we find on both mocks and real data no significant improvement from including non-Gaussian information, consistent with previous forecast on mock data (Cheng et al. 2020). There has been some confusion in the literature about the tightening of S_8 constraint from non-Gaussian statistics. We find that almost all the reported tightening of S_8 in the literature can be explained by two reasons. One is the aforementioned mismatch of scales between the non-Gaussian and Gaussian statistics. Many non-Gaussian statistics rely on map smoothing to cut scales, but a Gaussian smoothing kernel does not draw a sharp frequency cut (Figure 2). The second reason is leakage from the tightening of Ω_m due to remaining correlation between S_8 and Ω_m, which can be eliminated by optimizing the power index a in the definition of $S_8 \equiv \sigma_8(\Omega_m/0.3)^a$ or by fixing (conditioning on) Ω_m. Therefore, we conclude that there is no evidence that non-Gaussian statistics can significantly tighten the error bar of S_8. Nevertheless, non-Gaussian statistics do provide consistency checks for the S_8 constraint.
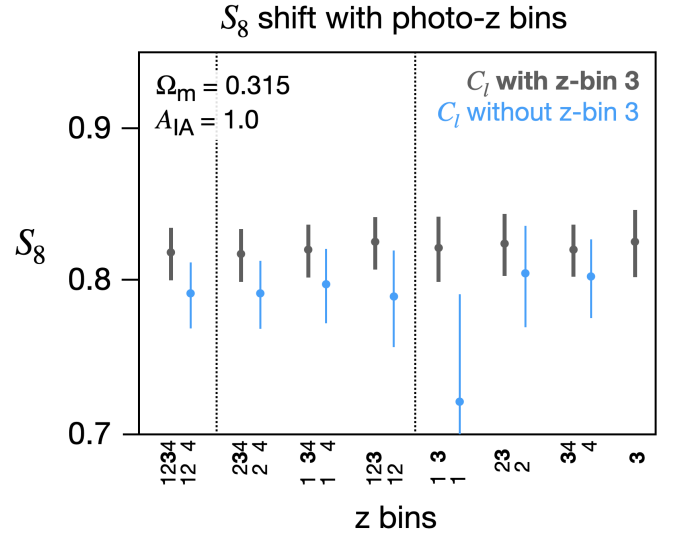
Figure 5. Shift of S_8 inference when using different subsets of redshift bins. The constraints are performed using tomographic wavelet power spectra (including the cross-spectra). To avoid error leaked from the poorly constrained Ω_m and A_{IA}, the inference is conditioned on $\Omega_m=0.315$ and $A_{IA}=0$. As long as the 3rd redshift bin (photo-z = [0.9-1.2]) is included, the S_8 value is stable around 0.82 (gray), otherwise it is around 0.79 (blue), indicating a potential underestimate of redshift in the 3rd bin.

3.2 Influence of photometric redshift on S_8

In this section we first point out an internal tension of S_8 estimates between different redshift bins, suggesting that photometric redshift is a main limiting factor for an accurate estimate of S_8. We then show that when adopting a similar treatment of photometric redshift, our result is consistent with other major results of HSC data analyses. However, changing the treatment of photometric redshift will substantially change the inferred value of S_8.

Figure 5 shows an internal consistency check of photo-z calibration using the inference from different subsets of tomographic maps. For example, the label '12 4' means that all the 6 auto- and cross-power spectra without bin 3, where bin 1–4 correspond to photo-z ranges of [0.3–0.6], [0.6–0.9], [0.9–1.2], and [1.2–1.5]. A systematic trend is found that as long as the 3rd redshift bin is absent, the inferred S_8 value is lower than those with bin 3 by 0.03. Similar patterns are also found when using the scattering transform or binned power spectra as summary statistics, in addition to the wavelet power spectra inference shown in Figure 5, so this pattern should not be caused by particular statistics but rather linked to the galaxy catalog or map itself.

Though such a lowering of S_8 (1.5σ) may be caused by statistical fluctuations, it may also be caused by an underestimate of the density of high redshift sources, i.e. a tail in the redshift distribution of the 3rd photo-z bin. Indeed, there are multiple challenges to obtaining reliable photometric redshift estimates when going to deeper magnitude and higher redshifts (e.g. Salvato et al. 2019). It requires modeling not only the spectra of galaxies within the redshift range of interest (0.3–1.5 in our case) but all galaxies, and the modeling error for any other galaxies can cause contamination into the redshift range of interest. The sensitivity comes from sharp features, mainly the Balmer and Lyman alpha breaks, in galaxy spectra. However, the narrow range of optical bands and photometric error lead to degenerate photo-z solutions. As a result, the underlying true redshift distribution of galaxies selected within our photo-z bins tends to have a long tail towards high redshifts, making it difficult to robustly esti-

---

[4] When compared to the power spectra with 6 bins between $l = 300$–1900, as adopted by Hikage et al. (2019), the improvement is a factor of 3, but apparently, the additional factor of 1.5 comes from extending the frequency range rather than including non-Gaussian information.
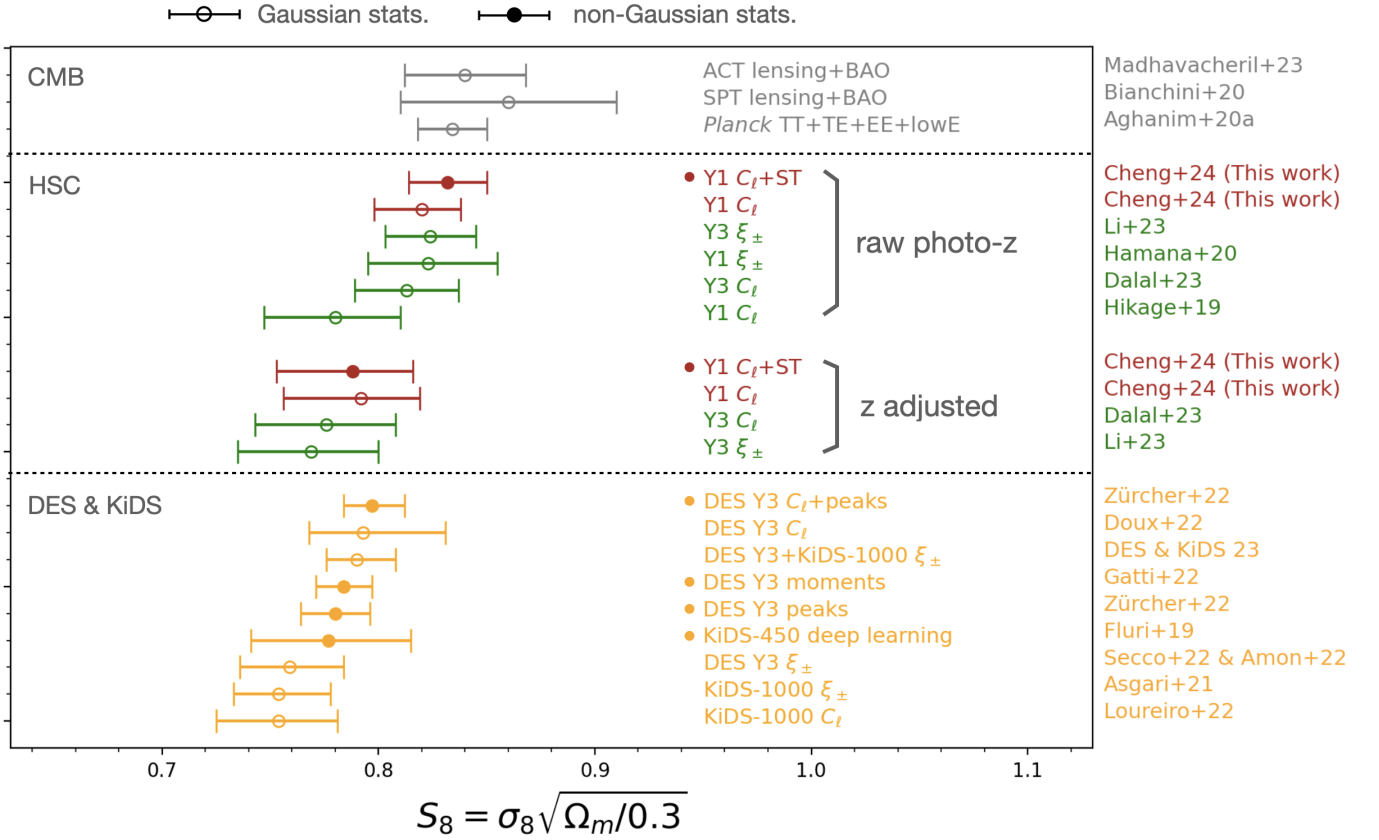
Figure 6. Comparison of S_8 inference. The upper part shows results from HSC, demonstrating the consistency when using the same raw photometric redshift and the shift towards lower values when adjusting the redshift: for our results, we removed bin 3 from the analysis; for HSC Y3, the authors set free biases of redshift calibration as nuisance parameters and joint constrain them by data. The lower part shows results from other stage-III weak lensing survey and CMB data: (Madhavacheril et al. 2023; Bianchini et al. 2020; Aghanim et al. 2020; Zürcher et al. 2022; Doux et al. 2022; DES and KiDS Collaboration 2023; Gatti et al. 2022; Fluri et al. 2019; Secco et al. 2022a; Amon et al. 2022; Asgari et al. 2021b; Loureiro et al. 2022).

mate the redshift distribution of the population. In addition, galaxies above $z = 2$ in general have high star formation rates and rich emission lines, making it difficult to build reliable templates for galaxy spectra.

In fact, a similar photo-z tension is also found for the 3rd and 4th redshift bins in the HSC Year 3 (Y3) data with 3 times more galaxies (Dalal et al. 2023; Li et al. 2023). As shown in Figure 6, the inferred S_8 values also cluster into two groups: with the raw photo-z used, values of $S_8 = 0.823^{+0.032}_{-0.028}$, 0.824 ± 0.021, 0.813 ± 0.024 are obtained by the correlation analysis of HSC Y1 data (Hamana et al. 2020), correlation analysis of HSC Y3 data (fig. 12 of Li et al. 2023), and pseudo-C_l analysis of HSC Y3 (Dalal et al. 2023). All of the three are consistent with each other and our result[5] at a 'high value' of S_8.

[5] When comparing their results to ours, we note that for photo-z prescription, they all use the COSMOS re-weighted redshift distribution from galaxies, while we use the stacked redshift posterior with the MLZ code. According to the tests presented in table 7 of Hikage et al. (2019) and figure 7 of Hamana et al. (2020), the inferred S_8 is higher by about 0.02 when switching from their method to ours. Therefore, before any special adjustment on photo-z, most HSC analyses, including our simulation-based inference, provide almost identical high values of S_8 at 0.82. The only analysis giving a low S_8 without photo-z adjustment is the pseudo-C_l analysis of HSC Y1 data reported in Hikage et al. (2019) ($S_8 = 0.78$). However, they also obtain an unusually

On the other hand, when a photo-z adjustment is performed, a similar lowering of S_8 is also found in the other HSC analyses. Instead of just excluding bin 3 as we did, the larger area of Y3 data allows for a simultaneous inference (self calibration) of photo-z errors. Dalal et al. (2023) and Li et al. (2023) assume that the shape of true redshift distribution in bin 3 and 4 are the same as that provided by the photo-z code, but an unknown overall offset is to be fitted. It turns out that the HSC Y3 data favors an uncomfortably large offset of $\Delta z \sim 0.1$ and 0.2 in the two tomographic bins, which dramatically lowers the inferred S_8 value: $S_8 = 0.769^{+0.031}_{-0.034}$ and 0.776 ± 0.032 are obtained from the correlation (Li et al. 2023) and power spectrum (Dalal et al. 2023) analyses, consistent with the lowering of S_8 we found in HSC Y1 data when excluding galaxies with photo-z between 0.9 and 1.2. Moreover, this internal tension is also observed with non-Gaussian statistics, including in our scattering transform analysis and the peak count analysis of the HSC Y1 data (fig. 9 of Marques et al. 2024a), despite the different choices of tomography and scale cut.

low $\Omega_m = 0.16$ not shared by the other analyses, which casts doubts on the reliability of their analysis.

3.3 Comparison to other S_8 estimates

Figure 6 shows the comparison of our S_8 inference with others in the literature. Regardless of which statistic is used, our results based on the raw redshift estimate from the HSC pipeline (Tanaka et al. 2018) have a 'high' value of S_8 consistent with *Planck*. As discussed in section 3.2, if we remove background sources with photometric redshifts in the 0.9–1.2 range (based on their "best point estimate" using the `Ephor-AB` code), the S_8 estimates become relatively lower and more consistent with those obtained with the KiDS[6] and DES[7] weak lensing surveys. Because of this internal redshift tension, it is currently difficult to robustly conclude on which value of S_8 is correct.

For a deep survey like the HSC, a significant fraction of galaxies have relatively high redshifts. Therefore, simply removing them from the analysis causes substantial loss of constraining power. The self-calibration method adopted by Dalal et al. (2023) and Li et al. (2023) is a practical solution but still not ideal, as the uncertainty of the redshift distribution is likely from the tail part which is not well captured by the simple overall shift. To definitively identify the nature of this internal tension from photo-z estimate, it may be essential to obtain independent redshift estimates, for example using the clustering redshift technique (e.g., Schneider et al. 2006; Newman 2008; Ménard et al. 2013; Schmidt et al. 2013; Chiang et al. 2019).

The above internal tension of S_8 suggests that photo-z calibration is a main limiting factor for S_8 inference in HSC and other ground based surveys reaching a similar depth, including the Rubin Observatory LSST[8] (LSST Science Collaboration et al. 2009). Before robust photo-z estimates become available, improvement of S_8 is unlikely to come from optimizing the choice of spatial statistic or machine learning tools [9]. As the photo-z bias is likely caused by the ambiguity of spectral structures such as the Balmer and Lyman alpha breaks, space-based surveys such as *Euclid*[10] (Laureijs et al. 2011), Xuntian space telescope survey(Zhan 2021), and *Roman* space telescope survey[11] (Spergel et al. 2015) which extend photometry to near infrared may significantly alleviate the photo-z problem and therefore draw conclusion on the S_8 tension.

4 CONCLUSION

We present the first application of the scattering transform to weak lensing observations. The scattering transform is a summary statistic that borrows ideas from convolutional neural nets, but requires no training and is compact and interpretable (see Cheng & Ménard 2021a for more details). To best demonstrate its power with real data, we choose to analyze the HSC Year 1 data (Mandelbaum et al. 2018a), which is the deepest wide-field weak lensing survey data available to date. From a full tomographic analysis of the HSC Y1 data, we conclude:

(i) *Spatial statistics:* We find that the scattering transform significantly improves the constraint on Ω_m. It tightens the error bar by a factor of 2 when compared to Gaussian statistics in the same scale range, reaching the constraining power of convolutional neural nets

(Lu et al. 2023). We also find that to improve Ω_m constraint, the non-Gaussian information across tomographic maps cannot be ignored, and it can be captured by measuring non-Gaussian statistics on linear combinations of different tomographic maps.

(ii) *Photo-z calibration:* We find that the inference of S_8 is more limited by photo-z calibration than choosing spatial statistics. When excluding galaxies with photo-z in [0.9-1.2] (the 3rd bin in HSC) from the analysis, the inferred S_8 changes from 0.82 to 0.79 independent of which statistic is used, suggesting an under-estimation of redshift of those galaxies. A similar shift is also found in the analyses of HSC Y3 data (Li et al. 2023; Dalal et al. 2023). The S_8 estimation is currently limited by photometric redshift estimation. This photo-z problem is likely shared by all ground-based weak lensing surveys reaching a similar depth. Calibration with clustering redshift and infrared photometry from space-based telescopes may help to overcome this problem and provide final answer to the S_8 tension.

DATA AVAILABILITY

The derived data generated in this research will be shared on reasonable request to the corresponding author.

REFERENCES

Abbott T. M. C., et al., 2022, Phys. Rev. D, 105, 023520
Aghanim N., et al., 2020, Astronomy & Astrophysics, 641, A8
Aihara H., et al., 2018, PASJ, 70, S4
Allys E., Levrier F., Zhang S., Colling C., Regaldo-Saint Blancard B., Boulanger F., Hennebelle P., Mallat S., 2019, A&A, 629, A115

Allys E., Marchand T., Cardoso J. F., Villaescusa-Navarro F., Ho S., Mallat S., 2020, arXiv e-prints, p. arXiv:2006.06298

Amon A., et al., 2022, Phys. Rev. D, 105, 023514

Anbajagane D., et al., 2023, MNRAS, 526, 5530

Asgari M., et al., 2021a, A&A, 645, A104

Asgari M., et al., 2021b, A&A, 645, A104

Bernstein G. M., Jarvis M., 2002, AJ, 123, 583

Bianchini F., et al., 2020, The Astrophysical Journal, 888, 119

Bruna J., Mallat S., 2013, IEEE Transactions on Pattern Analysis and Machine Intelligence, 35, 1872

Burger P. A., et al., 2023, arXiv e-prints, p. arXiv:2309.08602

Cheng S., Ménard B., 2021a, arXiv e-prints, p. arXiv:2112.01288

Cheng S., Ménard B., 2021b, MNRAS, 507, 1012

Cheng S., Ting Y.-S., Ménard B., Bruna J., 2020, MNRAS, 499, 5902

Cheng S., Morel R., Allys E., Ménard B., Mallat S., 2023, arXiv e-prints, p. arXiv:2306.17210

Chiang Y.-K., Ménard B., Schiminovich D., 2019, ApJ, 877, 150

DES and KiDS Collaboration 2023, arXiv e-prints, p. arXiv:2305.17173

Dai B., Seljak U., 2022, MNRAS, 516, 2363

Dai B., Seljak U., 2024, Proceedings of the National Academy of Science, 121, e2309624121

Dalal R., et al., 2023, Phys. Rev. D, 108, 123519

Doux C., et al., 2022, Monthly Notices of the Royal Astronomical Society, 515, 1942

Fluri J., Kacprzak T., Lucchi A., Refregier A., Amara A., Hofmann T., Schneider A., 2019, Phys. Rev. D, 100, 063514

Fluri J., Kacprzak T., Lucchi A., Schneider A., Refregier A., Hofmann T., 2022, Phys. Rev. D, 105, 083518

García-García C., Ruiz-Zapatero J., Alonso D., Bellini E., Ferreira P. G., Mueller E.-M., Nicola A., Ruiz-Lapuente P., 2021, J. Cosmology Astropart. Phys., 2021, 030

Gatti M., et al., 2022, Phys. Rev. D, 106, 083509

Gatti M., et al., 2023, arXiv e-prints, p. arXiv:2310.17557

Gatti M., et al., 2024, MNRAS, 527, L115

Grandón D., Marques G. A., Thiele L., Cheng S., Shirasaki M., Liu J., 2024, arXiv e-prints, p. arXiv:2403.03807

Hamana T., et al., 2020, PASJ, 72, 16

Harnois-Déraps J., Martinet N., Reischke R., 2022, MNRAS, 509, 3868

Heydenreich S., Brück B., Burger P., Harnois-Déraps J., Unruh S., Castro T., Dolag K., Martinet N., 2022, A&A, 667, A125

Hikage C., et al., 2019, PASJ, 71, 43

Hildebrandt H., et al., 2020, A&A, 633, A69

Hinshaw G., et al., 2013, ApJS, 208, 19

Hirata C., Seljak U., 2003, MNRAS, 343, 459

Hirata C. M., Seljak U., 2004, Phys. Rev. D, 70, 063526

Kaiser N., Squires G., 1993, ApJ, 404, 441

Kilbinger M., 2015, Reports on Progress in Physics, 78, 086901

LSST Science Collaboration et al., 2009, arXiv e-prints, p. arXiv:0912.0201

Laureijs R., et al., 2011, arXiv e-prints, p. arXiv:1110.3193

Li X., et al., 2023, Phys. Rev. D, 108, 123518

Liu X., Yuan S., Pan C., Zhang T., Wang Q., Fan Z., 2023, MNRAS, 519, 594

Loureiro A., et al., 2022, A&A, 665, A56

Lu T., Haiman Z., Li X., 2023, MNRAS, 521, 2050

Madhavacheril M. S., et al., 2023, arXiv preprint arXiv:2304.05203

Mandelbaum R., 2018, ARA&A, 56, 393

Mandelbaum R., et al., 2018a, PASJ, 70, S25

Mandelbaum R., et al., 2018b, MNRAS, 481, 3170

Marques G. A., et al., 2024a, MNRAS, 528, 4513

Marques G. A., et al., 2024b, J. Cosmology Astropart. Phys., 2024, 033

Martinet N., et al., 2018, MNRAS, 474, 712

Martinet N., Harnois-Déraps J., Jullo E., Schneider P., 2021, A&A, 646, A62

Ménard B., Scranton R., Schmidt S., Morrison C., Jeong D., Budavari T., Rahman M., 2013, arXiv e-prints, p. arXiv:1303.4722

Miralda-Escude J., 1991, ApJ, 380, 1

Nelson D., et al., 2019, Computational Astrophysics and Cosmology, 6, 2

Newman J. A., 2008, ApJ, 684, 88

Oguri M., et al., 2018, PASJ, 70, S26

Osato K., Liu J., Haiman Z., 2021, MNRAS, 502, 5593

Pires S., Starck J. L., Amara A., Teyssier R., Réfrégier A., Fadili J., 2009, MNRAS, 395, 1265

Planck Collaboration et al., 2016, A&A, 594, A13

Planck Collaboration et al., 2020, A&A, 641, A6

Ribli D., Pataki B. Á., Zorrilla Matilla J. M., Hsu D., Haiman Z., Csabai I., 2019, MNRAS, 490, 1843

Salvato M., Ilbert O., Hoyle B., 2019, Nature Astronomy, 3, 212

Schmidt S. J., Ménard B., Scranton R., Morrison C., McBride C. K., 2013, MNRAS, 431, 3307

Schneider M., Knox L., Zhan H., Connolly A., 2006, ApJ, 651, 14

Secco L. F., et al., 2022a, Phys. Rev. D, 105, 023515

Secco L. F., et al., 2022b, Phys. Rev. D, 105, 103537

Shirasaki M., Hamana T., Takada M., Takahashi R., Miyatake H., 2019, MNRAS, 486, 52

Shirasaki M., Moriwaki K., Oogi T., Yoshida N., Ikeda S., Nishimichi T., 2021, MNRAS, 504, 1825

Spergel D., et al., 2015, arXiv e-prints, p. arXiv:1503.03757

Takahashi R., Hamana T., Shirasaki M., Namikawa T., Nishimichi T., Osato K., Shiroyama K., 2017, ApJ, 850, 24

Tanaka M., et al., 2018, PASJ, 70, S9

Thiele L., Marques G. A., Liu J., Shirasaki M., 2023, Phys. Rev. D, 108, 123526

Valogiannis G., Dvorkin C., 2022, Phys. Rev. D, 106, 103509

Valogiannis G., Yuan S., Dvorkin C., 2023, arXiv e-prints, p. arXiv:2310.16116

Zhan H., 2021, Chinese Science Bulletin, 66, 1290

Zürcher D., et al., 2022, MNRAS, 511, 2075

APPENDIX A: SMOOTHING OF MASS MAP

With smoothing, the weighted average should be extended from within each pixel to galaxies in its neighborhood, therefore the weighting becomes a combination of the weights of individual galaxies, w_i and a smoothing kernel $W(\Delta r)$. Approximating the positions of each galaxy by the center of the pixels where they are located, a smooth $\bar{\phi}$ can be obtained by smoothing individually the numerator maps and the denominator map in eq. 2,

$$\phi(x, y)_{\text{smooth}} \equiv \frac{W(\Delta r) * \sum_{\text{in pixel (x,y)}} w_i \phi_i}{W(\Delta r) * \sum_{\text{in pixel (x,y)}} w_i} , \tag{A1}$$

where '$*$' represents convolution on a grid, and the smoothed fields $\phi(x, y)_{\text{smooth}}$ for the four quantities $\{e, \mathcal{R}, m_{\text{tot}}, c\}$ are used to generate the $\hat{\gamma}$ map through eq. 1. We use the Gaussian smoothing kernel with standard deviation $\sigma_r = 2$ arcmin:

$$W(\Delta r) = \frac{1}{2\pi\sigma_r^2} \exp\left(-\frac{\Delta r^2}{2\sigma_r^2}\right). \tag{A2}$$

In the literature, as little variation is expected in the map of m_{tot} and $\mathcal{R}$ in eq. 1, they are usually treated as constants over a field of view and redshift bin. We have confirmed that replacing $\bar{m}_{\text{tot}}$ and $\bar{\mathcal{R}}$ maps by the global weighted averages over all galaxies only results in negligible change in the cosmological inference. Therefore, our map making method is essentially the same as Oguri et al. (2018); Lu et al. (2023).

Below we discuss the advantage of this smoothing method compared to a direct convolution of $\bar{\phi}(x, y)$, as adopted in Thiele et al. (2023); Marques et al. (2024a). Compared to eq. A1 where both the numerator and denominator are convolved by the smoothing kernel, an alternative is to directly convolve the pixelized map by the smoothing kernel,

$$\phi'(x, y)_{\text{smooth}} \equiv W(\Delta r) * \frac{\sum_{\text{in pixel (x,y)}} w_i \phi_i}{\sum_{\text{in pixel (x,y)}} w_i} , \tag{A3}$$

However, such a direct convolution is not an optimal estimator because it weighs pixels equally while the effective number of galaxies in each pixel is different. Thus it increases the noise level of the map and reduces the constraining power on cosmological parameters. This is a main difference between the maps used in Marques et al. (2024a); Thiele et al. (2023) and this paper. To give a quantitative comparison, on the non-tomographic mass map at fiducial cosmology smoothed by a 1 arcmin kernel, our weighted smoothing method provides a correlation coefficient of 0.38 (an S/N of 0.37) between a noisy mock and the noiseless convergence map, whereas the equal-weight smoothing gives a correlation coefficient of 0.32 (an S/N of 0.22). The error bar of inference increases by 50% in Ω_m and 25% in $S_8 \equiv \sigma_8(\Omega_m/0.3)^{0.5}$ on both mock and real data. In power spectrum and correlation function analyses (e.g., Hikage et al. 2019; Hamana et al. 2020), although there is no explicit smoothing in the map making step, a treatment of weighting equivalent to eq. A1 is included in the formalism of computing their statistics.

APPENDIX B: MORLET WAVELET

The profile of the Morlet wavelets we use is given by

$$\psi^{j,\theta}(\vec{x}) = \frac{1}{\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) [\exp(i\vec{k}_0 \cdot \vec{x}) - \beta], \tag{B1}$$

$$\tilde{\psi}^{j,\theta}(\vec{k}) = \exp\left[-\frac{(\vec{k} - \vec{k}_0)^2\sigma^2}{2}\right] - \beta \exp\left(-\frac{k^2\sigma^2}{2}\right), \tag{B2}$$

where $\tilde{}$ represents Fourier transform, and

$$\sigma = 0.8 \times 2^j, \ |k_0| = 0.75\pi \times 2^{-j}, \ \text{angle}(k_0) = \pi\theta/L \tag{B3}$$

specify the dependence on wavelet size index j and orientation index θ. The offset $\beta = \exp\left(-\frac{k_0^2\sigma^2}{2}\right)$ is introduced to guarantee that the wavelets are strictly band-pass instead of low-pass filters: $\tilde{\psi}(0) = 0$. More explanations of the mathematical form can be found in the appendix B of Cheng & Ménard (2021a).

As a result, the central frequency of the $j = 0$ wavelet kernel is defined to be 3/8 pixel^{-1}, and when j increases by 1, the wavelet becomes larger in real space by a factor of 2. In this study we use wavelets with $j = 2$–4. The $j = 2$ wavelet roughly covers $l \approx 1200$–3500, $j = 3$ wavelet covers $l \approx 600$–1800, and $j = 4$ wavelet covers $l \approx 300$–900. In principle the sampling of j does not have to be integers, nevertheless we still sample j at integers as we find no significant improvement of cosmological constraint from finer sampling. As for orientations, we follow previous studies and set the total number of orientations to be $L = 4$, i.e., sampling every 45 degrees.

APPENDIX C: SIMULATION BASED INFERENCE METHOD

C1 Mock catalogs

The mock catalogs are generated following Shirasaki et al. (2019) by first randomly rotating the observed galaxies and statistically separating the measurement noise from intrinsic galaxy shape,

$$e^{\text{int}} = \text{random rotate}\left(\frac{e_{\text{rms}}}{\sqrt{e_{\text{rms}}^2 + \sigma_e^2}}\right) e^{\text{obs}} \tag{C1}$$

$$e^{\text{mea}} = \sigma_e N \tag{C2}$$

where e_{rms} and σ_e are read from the observed galaxy catalog (Mandelbaum et al. 2018b) and N is a random variable independently sampled for each galaxy from a standard Gaussian distribution with zero mean and unity variance Then, the mock ellipticity catalog is obtained by applying cosmic shear distortion δ^{sh} to the intrinsic galaxy shapes and then adding mock measurement noise (Miralda-Escude 1991; Bernstein & Jarvis 2002)[12]

$$e_1 = \frac{e_1^{\text{int}} + \delta_1^{\text{sh}} + a\delta_2^{\text{sh}}(\delta_1^{\text{sh}}e_2^{\text{int}} - \delta_2^{\text{sh}}e_1^{\text{int}})}{1 + \delta^{\text{sh}} \cdot e^{\text{int}}} + e_1^{\text{mea}} \tag{C3}$$

$$e_2 = \frac{e_2^{\text{int}} + \delta_2^{\text{sh}} + a\delta_1^{\text{sh}}(\delta_2^{\text{sh}}e_1^{\text{int}} - \delta_1^{\text{sh}}e_2^{\text{int}})}{1 + \delta^{\text{sh}} \cdot e^{\text{int}}} + e_2^{\text{mea}} \tag{C4}$$

$$a \equiv \frac{1 - \sqrt{1 - (\delta^{\text{sh}})^2}}{(\delta^{\text{sh}})^2} = \frac{1}{1 + \sqrt{1 - (\delta^{\text{sh}})^2}} \tag{C5}$$

$$\delta^{\text{sh}} \equiv \frac{2(1 - \kappa)}{(1 - \kappa)^2 + \gamma^2}\gamma, \tag{C6}$$

where a is a factor close to 0.5 at the small shear limit, and the shear distortion δ^{sh} is obtained by ray-tracing the N-body simulations. We also simulate the effect of multiplicative bias m_{tot} by replacing γ with $(1 + m_{\text{tot}})\gamma$.

The fiducial simulation set (Takahashi et al. 2017; Shirasaki et al. 2019) includes 108 quasi-independent full-sky realizations at one fiducial cosmology with the ΛCDM model and WMAP9 parameters (Hinshaw et al. 2013): $\Omega_{\text{cdm}} = 0.233$, $\Omega_b = 0.046$, $\Omega_m = \Omega_{\text{cdm}} + \Omega_b = 0.279$, $\Omega_\Lambda = 0.721$, $h = 0.7$, $\sigma_8 = 0.82$, and $n_s = 0.97$. The sky footprint of the HSC fields was rotated in 21 angles to take full advantage of the large volume of each simulation, resulting in a total of 2,268 realizations of mock HSC lensing maps. More details are described in Appendix D and the references. This set of mock data at the fiducial cosmology is used to estimate the cosmic variance.

The cosmic-varying simulation set in (Shirasaki et al. 2021) samples 100 different cosmologies with $0.1 < \Omega_m < 0.7$ and $0.23 < S_8 < 1.1$. The other parameters are consistent with the *Planck* 2015 results (Planck Collaboration et al. 2016): $\Omega_b = 0.049$, $\Omega_m + \Omega_\Lambda = 1$, $h = 0.6727$, and $n_s = 0.9645$. We make 50 realizations per cosmology by random arrangement of the simulation boxes and further enlarge the size to 150 by adding different realizations of galaxy noise.

To obtain the κ and γ in eq. C6 to enable generation of the mock galaxy shape catalog, one needs to assign a redshift to each of the galaxies. We follow the same method detailed in (Shirasaki et al. 2019). In brief, we use the stacked redshift posterior of galaxies provided by the MLZ code, a photo-z code using a self-organizing map, in each tomographic bin as the assumed redshift distribution.

C2 Likelihood emulator

We parameterize the likelihood function by assuming that at each cosmology, the summary statistics $\vec{s}$ follow a multivariate Gaussian distribution under cosmic variance and galaxy shape noise:

$$p(\vec{s}|\theta) = \frac{\exp\left[-\frac{1}{2}(\vec{s} - \vec{\mu}(\theta))^T \mathbf{C}_{\text{fid}}^{-1}(\vec{s} - \vec{\mu}(\theta))\right]}{(2\pi|\mathbf{C}_{\text{fid}}|)^{\frac{\dim(\vec{s})}{2}}}. \tag{C7}$$

---

[12] With some algebra, one can rewrite these equations in a compact form, which can be found in the literature, in terms of the complex-valued reduced shear $g \equiv \frac{\gamma}{1-\kappa}$ and complex ellipticity: $e = \frac{e^{\text{int}} + 2g + g^2 e^{\text{int}*}}{1 + |g|^2 + 2\text{Re}(ge^{\text{int}*})}$, by using the relation $\delta^{\text{sh}} = \frac{2g}{1+|g|^2}$ and $a = \frac{1+|g|^2}{2}$.

Usually, this assumption is justified by the fact that the statistical estimator is an average over a large field of view, for which the central limit theorem Gaussianizes its distribution. This assumption is particularly well justified for the scattering transform coefficients because the modulus operation avoids amplification of outliers and thus accelerates the Gaussianization (Cheng & Ménard 2021b).

We estimate the mean $\mu(\theta)$ as a function of cosmological parameters $\theta = (\Omega_m, S_8)$ from the set of cosmic varied simulations, by first calculating the sample mean at each cosmology $\hat{\mu}(\theta_i)$ and then fit it with 6th-order polynomials of $(\log \Omega_m, \log S_8)$.

$$\vec{\mu}(\Omega_{\mathrm{m}}, S_8) = \sum_{0 \le n+m \le 6} \vec{a}_{n,m} (\log \Omega_{\mathrm{m}})^n (\log S_8)^m , \qquad (C8)$$

where n and m are natural numbers. The reason to use logarithms is that $\mu(\Omega_{\mathrm{m}}, S_8)$ or its derivatives may have singularity at $\Omega_{\mathrm{m}} = 0$ or $S_8 = 0$. As the summary statistics may change quickly when Ω_{m} approaches zero, we use a logarithm to regularize this behavior and make the polynomial regression converge faster. The regression is implemented through least square method for each component of $\vec{a}_{n,m}$. We verify the robustness of our emulator by fitting it with 5th-order and 7th-order polynomials and find no significant difference, which means our choice does not under-fit or over-fit. The covariance matrix $\mathbf{C}_{\mathrm{fid}}$ in our model is fixed and does not change with cosmological parameters. We estimate it only with the 2,268 fiducial realizations that originate from 108 quasi-independent simulations and a full-sky setting. In addition, we find small differences between the fiducial mocks and the emulator prediction based on the cosmo-varied simulations, which is likely caused by the small and adaptive box sizes used in the cosmo-varied simulations. We calibrate the emulator using the mean value of fiducial mocks, which is described in detail in appendix D.

C3 Systematics

To account for observational and astrophysical systematics, we include the following parameters in our emulator: multiplicative calibration error Δm, galaxy intrinsic alignment strength A_{IA}, and baryonic feedback strength A_{tng}. We further model their influence to the summary statistics with a simple form,

$$\vec{\mu}(\Omega_{\mathrm{m}}, S_8, A_{\mathrm{IA}}, \Delta m, A_{\mathrm{tng}}) = \vec{b}^{A_{\mathrm{IA}}} \vec{c}^{A_{\mathrm{tng}}} \vec{d}^{\Delta m} \vec{\mu}(\Omega_{\mathrm{m}}, S_8, 0, 0, 0) , \qquad (C9)$$

where the coefficients $\vec{b}$, $\vec{c}$, $\vec{d}$ are constants estimated using simulations described below and the operations are meant as element-wise. That parameterization is equivalent to a linear model for the logarithm of our statistics,

$$\log \vec{\mu}(\Omega_{\mathrm{m}}, S_8, A_{\mathrm{IA}}, \Delta m, A_{\mathrm{tng}}) =$$
$$A_{\mathrm{IA}} \log \vec{b} + A_{\mathrm{tng}} \log \vec{c} + \Delta m \log \vec{d} + \log \vec{\mu}(\Omega_{\mathrm{m}}, S_8, 0, 0, 0) . \qquad (C10)$$

We consider it better than a linear model of the original statistics because both the scattering coefficients and the power spectrum are always positive and related to fluctuation strength. This relation is accurate in the neighborhood of the fiducial cosmology, though its accuracy may gradually degrade as one moves away from the fiducial.

C3.1 Intrinsic alignment

To incorporate intrinsic alignment to our emulator for non-Gaussian statistics, we use a mock catalog provide in Harnois-Déraps et al. (2022) which simulates this effect. This catalog was made following

a prescription called δ-NLA, which computes the intrinsic alignment field e^{IA} according to the non-linear tidal alignment model (NLA) (Hirata & Seljak 2004) but also samples galaxies naturally according to the matter density field δ. Therefore, compared to the NLA model, it realistically up-weights regions with higher matter density. The intrinsic alignment strength is still parametrized by A_{IA} as in NLA. For simplicity we assume no evolution of A_{IA} with redshift. Previous studies with HSC Y1 data obtains consistent constraints with different error bar sizes, $A_{\mathrm{IA}} = 0.4 \pm 0.7$ (Hikage et al. 2019) and $A_{\mathrm{IA}} = 0.9 \pm 0.3$ (Hamana et al. 2020), with the NLA model. According to those approximate range of A_{IA}, we build mocks with $A_{\mathrm{IA}} = 0$ and 1.5, and determine the coefficient $\vec{b}$ in eq. C9 by $\vec{b} = \left(\frac{\langle \vec{s} \rangle_{A_{\mathrm{IA}}=1.5}}{\langle \vec{s} \rangle_{A_{\mathrm{IA}}=0}} \right)^{\frac{1}{1.5}}$.

The average is measured from 400 mocks with $10 \times 10 \ \mathrm{deg}^2$ field of view, generated by adding different galaxy shape noise realizations to the 10 mock catalogs provided in Harnois-Déraps et al. (2022).

C3.2 Error in multiplicative bias

To account for the effect of mis-calibration of multiplicative bias, we model its influence on summary statistics and then marginalize over it. A prior of a Gaussian with 0.01 standard deviation is applied, which is the level of calibration residual estimated from image simulations (Mandelbaum et al. 2018b). We determine the coefficient c in eq. C9 by its value at fiducial cosmology $\vec{d} = \left(\frac{\langle \vec{s} \rangle_{\mathrm{fid}, \Delta m=+0.01}}{\langle \vec{s} \rangle_{\mathrm{fid}, \Delta m=-0.01}} \right)^{\frac{1}{0.02}}$, which are obtained by adding $\Delta m = +0.01$ and -0.01 to the multiplicative bias of mock galaxies in the 2,268 realizations of fiducial mock catalogs, while keeping all other ingredients including the random seed for galaxy shape noise unchanged. We find that the effect of multiplicative bias error basically shifts S_8 by the same fraction, while not changing the inference of Ω_{m}. Data itself is not able to constrain Δm and it is mainly driven by the prior.

C3.3 Baryonic feedback

We use the kappa-TNG mocks (Osato et al. 2021) to test baryonic effects. They are a set of convergence maps made from the hydrodynamic IllustrisTNG simulations (Nelson et al. 2019), with 10,000 realizations of $5 \times 5 \ \mathrm{deg}^2$ maps ray-traced from the highest resolution of the largest box TNG300-1 and its corresponding dark-matter-only counterpart TNG300-1-Dark. Those simulations adopts a cosmology consistent with Planck Collaboration et al. (2016). Since these mocks are not tailored for the HSC Y1 field of view, the map statistics are not defined identically to the emulator or observation due to mask and edge effects. The coefficient $\vec{c}$ in eq. C9 is then determined by the ratio between hydro and dark matter only simulations: $\vec{c} = \frac{\langle \vec{s} \rangle_{\mathrm{TNG}}}{\langle \vec{s} \rangle_{\mathrm{TNG\text{-}dark}}}$.

C3.4 Photometric redshift algorithms

To test the influence of photometric redshift error, we generate 210 new mocks at the fiducial cosmology using photo-z from two other codes, `frankenz` and `mizuki` (Tanaka et al. 2018). Due to the limited number of additional mocks, we are not able to incorporate Δz_{phot} as a nuisance parameter and marginalize over it. Therefore, we only apply the ratio $\langle \vec{s} \rangle_{\mathrm{fid}, \texttt{frankenz}} / \langle \vec{s} \rangle_{\mathrm{fid}, \texttt{MLZ}}$ and $\langle \vec{s} \rangle_{\mathrm{fid}, \texttt{mizuki}} / \langle \vec{s} \rangle_{\mathrm{fid}, \texttt{MLZ}}$ to the emulator and examine the shift of cosmological inference.

C3.5 PSF residual

We also include PSF systematics in our model. Errors in the measurement of point spread function (PSF) and the PSF leakage from imperfect deconvolution lead to additive bias of the shear measurement. The residual of this calibration acts like correlated noises added to the shear map. To account for its effect, we simulate this noise in our mock maps by adding a Gaussian random field to the mock convergence maps, with the four redshift bins sharing the same PSF systematics. We set the power spectrum of this noise using Figure 7 of Lu et al. (2023), which corresponds to setting $\alpha_{\mathrm{psf}} = 0.03$ and $\beta_{\mathrm{psf}} = -0.89$ as measured from data by Hamana et al. (2020). For simplicity we omit the uncertainty of this power spectrum. Assuming no PSF residual typically shifts S_8 up by 0.004 and Ω_m up by 0.01. A more complicated treatment such as a marginalization of the parametrized power spectrum of the PSF systematics is left for future work.

C3.6 Source clustering

We have not modeled the effect of source clustering, but according to an exploration performed by Gatti et al. (2023), we expect negligible biased for the scattering transform and power spectrum. Source clustering is the fact that background galaxies are not uniformly distributed but clustered with some correlation of the foreground. Therefore, the observed lensing map is sampled in a biased way and have different statistical properties from uniformly sampled ones as used in our simulations. Nevertheless, source clustering mostly influences statistics that are odd function of the field (i.e., statistics that change their sign when the field is flipped) (e.g., Gatti et al. 2024). As a result, the power spectrum and the scattering transform are not much influenced but source clustering, whereas third-order moment based statistics, phase harmonic coefficients that are odd, and machine-learning methods may be significantly biased if the source clustering effect is not properly modeled.

APPENDIX D: *N*-BODY SIMULATIONS

The fiducial simulation set (Takahashi et al. 2017; Shirasaki et al. 2019) includes 108 quasi-independent full-sky realizations at one fiducial cosmology with the ΛCDM model and WMAP9 parameters (Hinshaw et al. 2013): $\Omega_{\mathrm{cdm}} = 0.233$, $\Omega_b = 0.046$, $\Omega_m = \Omega_{\mathrm{cdm}} + \Omega_b = 0.279$, $\Omega_\Lambda = 0.721$, $h = 0.7$, $\sigma_8 = 0.82$, and $n_s = 0.97$. In total 6 independent runs, each with 14 simulations with different box sizes (450–$6300 h^{-1}$ Mpc), were implemented to keep the angular resolution roughly the same along the line of sight. Each box has 2048^3 particles corresponding to an angular particle grid size of 3.4–6.8 arcmin at the initial condition. The particles in each box are also randomly shifted and ray-traced to create 108 quasi-independent full sky maps with HEALPix resolution $N_{\mathrm{side}}=8192$. It has been tested that for lensing convergence with sources at $z = 1$, the power spectrum in simulation is consistent with halo-fit results within 5% till around $l = 3000$ (see Fig. 23 in Takahashi et al. (2017) and eq. 26 in Shirasaki et al. (2019)).

The cosmic-varying simulation set in (Shirasaki et al. 2021) samples 100 different cosmologies with $0.1 < \Omega_m < 0.7$ and $0.23 < S_8 < 1.1$. The other parameters are consistent with the *Planck* 2015 results (Planck Collaboration et al. 2016): $\Omega_b = 0.049$, $\Omega_m + \Omega_\Lambda = 1$, $h = 0.6727$, and $n_s = 0.9645$. At each cosmology, four *N*-body simulations with different box sizes and resolutions were run and connected along the light cone to cover a 10×10 deg^2 field

of view. The box sizes are designed to match the opening angle at all redshifts, with typical values at $\Omega_m = 0.3$ to be 280–930 Mpc/h. The projected particle grid at the initial condition in these simulations has a size of 1.3–2.5 arcmin, which is 2.5 times finer than the fiducial set of simulations. By randomly shifting each simulation box assuming periodic boundary condition, 50 cosmic shear realizations were generated for each cosmology. From stability tests such as using only a subset of the realizations to build the emulator, we find that 50 realizations are not enough to overcome the numerical noise of the cosmological dependence to mean data vector, so we enlarge the mocks to 150 realizations per cosmology by adding different galaxy noise.

We find that the $\mu(\theta)$ predicted by the emulator at the fiducial cosmology has an offset from the sample mean $\langle \vec{s} \rangle$ of the fiducial simulations. For all tomographic maps, our emulator based on cosmo-varied simulations predicts a lower power spectrum. The deviation is highest for low redshift and for large scales. The deviation in cosmic shear power spectrum (noise removed) is about 20% for the 1st tomographic bin and 10% for the 4th bin at $l = 300$. The deficit gradually changes to zero at around $l = 2500$ and becomes an excess after that. The excess is expected as the cosmo-varied simulations have higher resolution, but the cause of power deficit is unclear. This difference, dominated by the large scale power deficit, results in an offset in parameter inference. When using the mean of fiducial data vector as input for mock inference, we obtain a best fit of $\Omega_m = 0.261$ and $S_8 = 0.808$, 6.5% lower in Ω_m and 2% higher in S_8 than the fiducial values. We suspect that the deficit of power of our cosmo-varied maps based on Shirasaki et al. (2021) is caused by the adaptive box size of simulations along the light cone, as this is the main difference from another simulation based analysis of HSC year 1 data (Lu et al. 2023), where no such effect is observed. The adaptive box size results in the fact that a large angular scale always corresponds to roughly the simulation box size, at which the power may be suppressed due to spectral leakage under windowing. For this study, we just correct for this effect by multiplying the emulator prediction of $\vec{\mu}$ by a factor vector $\langle \vec{s} \rangle_{\mathrm{fid}} / \vec{\mu}(0.279, 0.791)_{\mathrm{emulator}}$. A full understanding of the deficit is left for future study.

This paper has been typeset from a TeX/LaTeX file prepared by the author.