PHYSICAL REVIEW D **113,** 083501 (2026)

Dark Energy Survey Year 3 results: Simulation-based *w*CDM inference from weak lensing and galaxy clustering maps with deep learning: Analysis design

A. Thomsen,[1,*] J. Bucko,[1] T. Kacprzak,[2,3] V. Ajani,[4] J. Fluri,[5] A. Refregier,[1] D. Anbajagane,[6] F. J. Castander,[7,8] A. Ferté,[9]
M. Gatti,[6] N. Jeffrey,[10] A. Alarcon,[8] A. Amon,[11] K. Bechtol,[12] M. R. Becker,[13] G. M. Bernstein,[14] A. Campos,[15,16]
A. Carnero Rosell,[17,18,19] C. Chang,[20,6] R. Chen,[21] A. Choi,[22] M. Crocce,[7,8] C. Davis,[23] J. DeRose,[24] S. Dodelson,[20,25,6]
C. Doux,[14,26] K. Eckert,[14] J. Elvin-Poole,[27] S. Everett,[28] P. Fosalba,[7,8] D. Gruen,[29] I. Harrison,[30] K. Herner,[25] E. M. Huff,[28,31]
M. Jarvis,[14] N. Kuropatkin,[25] P.-F. Leget,[23] N. MacCrann,[32] J. McCullough,[11,23,9,29] J. Myles,[11] A. Navarro-Alsina,[33]
S. Pandey,[14] A. Porredon,[34,35] J. Prat,[20,36] M. Raveri,[37] M. Rodriguez-Monroy,[38] R. P. Rollins,[39] A. Roodman,[23,9]
E. S. Rykoff,[23,9] C. Sánchez,[14] L. F. Secco,[6] E. Sheldon,[40] T. Shin,[41] M. A. Troxel,[21] I. Tutusaus,[42] T. N. Varga,[43,44,45]
N. Weaverdyck,[46,24] R. H. Wechsler,[47,23,9] B. Yanny,[25] B. Yin,[21] Y. Zhang,[48] J. Zuntz,[49] M. Aguena,[50,51] S. Allam,[25]
F. Andrade-Oliveira,[52] D. Bacon,[53] J. Blazek,[54] D. Brooks,[10] R. Camilleri,[55] J. Carretero,[56] R. Cawthon,[57] L. N. da Costa,[18]
M. E. da Silva Pereira,[58] T. M. Davis,[55] J. De Vicente,[34] S. Desai,[59] P. Doel,[10] J. García-Bellido,[60] G. Gutierrez,[25]
S. R. Hinton,[55] D. L. Hollowood,[61] K. Honscheid,[62,63] D. J. James,[64] K. Kuehn,[65,66] O. Lahav,[10] S. Lee,[31] J. L. Marshall,[67]
J. Mena-Fernández,[26] F. Menanteau,[68,69] R. Miquel,[70,56] J. Muir,[71,72] R. L. C. Ogando,[73] A. A. Plazas Malagón,[23,9]
E. Sanchez,[34] D. Sanchez Cid,[34,52] I. Sevilla-Noarbe,[34] M. Smith,[74] E. Suchyta,[75] M. E. C. Swanson,[68]
D. Thomas,[53] C. To,[20] and D. L. Tucker[25]

(DES Collaboration)

[1]*Department of Physics, ETH Zurich, Wolfgang-Pauli-Strasse 16, CH-8093 Zurich, Switzerland*
[2]*University Observatory Munich, Scheinerstraße 1, D-81679 Munich, Germany*
[3]*University of Applied Sciences Northwestern Switzerland FHNW,*
Bahnhofstrasse 6, 5210 Windisch, Switzerland
[4]*Fondazione LINKS, Via Pier Carlo Boggio 61, 10138 Turin, Italy*
[5]*Department of Computer Science, ETH Zurich, 8092 Zurich, Switzerland*
[6]*Kavli Institute for Cosmological Physics, University of Chicago, Chicago, Illinois 60637, USA*
[7]*Institut d'Estudis Espacials de Catalunya (IEEC), 08034 Barcelona, Spain*
[8]*Institute of Space Sciences (ICE, CSIC),*
Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain
[9]*SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA*
[10]*Department of Physics and Astronomy, University College London,*
Gower Street, London, WC1E 6BT, United Kingdom
[11]*Department of Astrophysical Sciences, Princeton University,*
Peyton Hall, Princeton, New Jersey 08544, USA
[12]*Physics Department, 2320 Chamberlin Hall, University of Wisconsin-Madison,*
1150 University Avenue Madison, Wisconsin 53706-1390, USA
[13]*Argonne National Laboratory, 9700 South Cass Avenue, Lemont, Illinois 60439, USA*
[14]*Department of Physics and Astronomy, University of Pennsylvania,*
Philadelphia, Pennsylvania 19104, USA
[15]*Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15312, USA*
[16]*NSF AI Planning Institute for Physics of the Future, Carnegie Mellon University,*
Pittsburgh, Pennsylvania 15213, USA
[17]*Instituto de Astrofisica de Canarias, E-38205 La Laguna, Tenerife, Spain*
[18]*Laboratório Interinstitucional de e-Astronomia - LIneA, Avenida Pastor Martin Luther King Jr,*
126 Del Castilho, Nova América Offices, Torre 3000/sala 817 CEP: 20765-000, Brazil
[19]*Universidad de La Laguna, Departamento Astrofísica, E-38206 La Laguna, Tenerife, Spain*
[20]*Department of Astronomy and Astrophysics, University of Chicago, Chicago, Illinois 60637, USA*
[21]*Department of Physics, Duke University, Durham, North Carolina 27708, USA*
[22]*NASA Goddard Space Flight Center, 8800 Greenbelt Road, Greenbelt, Maryland 20771, USA*
[23]*Kavli Institute for Particle Astrophysics and Cosmology, P.O. Box 2450, Stanford University,*
Stanford, California 94305, USA
[24]*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA*
[25]*Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, Illinois 60510, USA*
[26]*Université Grenoble Alpes, CNRS, LPSC-IN2P3, 38000 Grenoble, France*

2470-0010/2026/113(8)/083501(39)

Published by the American Physical Society

[27]*Department of Physics and Astronomy, University of Waterloo,*
200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada
[28]*California Institute of Technology, 1200 East California Boulevard, MC 249-17,*
Pasadena, California 91125, USA
[29]*University Observatory, LMU Faculty of Physics, Scheinerstrasse 1, 81679 Munich, Germany*
[30]*School of Physics and Astronomy, Cardiff University, CF24 3AA, United Kingdom*
[31]*Jet Propulsion Laboratory, California Institute of Technology,*
4800 Oak Grove Drive, Pasadena, California 91109, USA
[32]*Department of Applied Mathematics and Theoretical Physics, University of Cambridge,*
Cambridge CB3 0WA, United Kingdom
[33]*Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas,*
13083-859, Campinas, São Paulo, Brazil
[34]*Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain*
[35]*Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute,*
German Centre for Cosmological Lensing, 44780 Bochum, Germany
[36]*Nordita, KTH Royal Institute of Technology and Stockholm University,*
Hannes Alfvéns väg 12, SE-10691 Stockholm, Sweden
[37]*Department of Physics, University of Genova and INFN, Via Dodecaneso 33, 16146, Genova, Italy*
[38]*Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain*
[39]*Jodrell Bank Center for Astrophysics, School of Physics and Astronomy, University of Manchester,*
Oxford Road, Manchester M13 9PL, United Kingdom
[40]*Brookhaven National Laboratory, Building 510, Upton, New York 11973, USA*
[41]*Department of Physics and Astronomy, Stony Brook University, Stony Brook, New York 11794, USA*
[42]*Institut de Recherche en Astrophysique et Planétologie (IRAP), Université de Toulouse,*
CNRS, UPS, CNES, 14 Avenue Edouard Belin, 31400 Toulouse, France
[43]*Excellence Cluster Origins, Boltzmannstrasse 2, 85748 Garching, Germany*
[44]*Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, 85748 Garching, Germany*
[45]*Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München,*
Scheinerstrasse 1, 81679 München, Germany
[46]*Berkeley Center for Cosmological Physics, Department of Physics, University of California,*
Berkeley, California 94720, USA
[47]*Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, California 94305, USA*
[48]*Cerro Tololo Inter-American Observatory,*
NSF's National Optical-Infrared Astronomy Research Laboratory, Casilla 603, La Serena, Chile
[49]*Institute for Astronomy, University of Edinburgh, Edinburgh EH9 3HJ, United Kingdom*
[50]*Laboratório Interinstitucional de e-Astronomia - LIneA, Avenida Pastor Martin Luther King Jr,*
126 Del Castilho, Nova América Offices, Torre 3000/sala 817 CEP: 20765-000, Brazil
[51]*INAF-Osservatorio Astronomico di Trieste, via G. B. Tiepolo 11, I-34143 Trieste, Italy*
[52]*Physik-Institut, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland*
[53]*Institute of Cosmology and Gravitation, University of Portsmouth,*
Portsmouth, PO1 3FX, United Kingdom
[54]*Department of Physics, Northeastern University, Boston, Massachusetts 02115, USA*
[55]*School of Mathematics and Physics, University of Queensland, Brisbane, Queensland 4072, Australia*
[56]*Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology,*
Campus UAB, 08193 Bellaterra (Barcelona) Spain
[57]*Oxford College of Emory University, Oxford, Georgia 30054, USA*
[58]*Hamburger Sternwarte, Universität Hamburg, Gojenbergsweg 112, 21029 Hamburg, Germany*
[59]*Department of Physics, IIT Hyderabad, Kandi, Telangana 502285, India*
[60]*Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, 28049 Madrid, Spain*
[61]*Santa Cruz Institute for Particle Physics, Santa Cruz, California 95064, USA*
[62]*Center for Cosmology and Astro-Particle Physics, The Ohio State University,*
Columbus, Ohio 43210, USA
[63]*Department of Physics, The Ohio State University, Columbus, Ohio 43210, USA*
[64]*Center for Astrophysics | Harvard & Smithsonian,*
60 Garden Street, Cambridge, Massachusetts 02138, USA
[65]*Australian Astronomical Optics, Macquarie University, North Ryde, New South Wales 2113, Australia*
[66]*Lowell Observatory, 1400 Mars Hill Road, Flagstaff, Arizona 86001, USA*
[67]*George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, and*
Department of Physics and Astronomy, Texas A&M University, College Station, Texas 77843, USA

[68]*Center for Astrophysical Surveys, National Center for Supercomputing Applications,*
1205 West Clark Street, Urbana, Illinois 61801, USA
[69]*Department of Astronomy, University of Illinois at Urbana-Champaign,*
1002 West Green Street, Urbana, Illinois 61801, USA
[70]*Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain*
[71]*Department of Physics, University of Cincinnati, Cincinnati, Ohio 45221, USA*
[72]*Perimeter Institute for Theoretical Physics,*
31 Caroline Street North, Waterloo, Ontario N2L 2Y5, Canada
[73]*Centro de Tecnologia da Informação Renato Archer, Campinas, São Paulo,*
Brazil 13069-901, Observatório Nacional, Rio de Janeiro, RJ, Brazil 20921-400
[74]*Physics Department, Lancaster University, Lancaster, LA1 4YB, United Kingdom*
[75]*Computer Science and Mathematics Division, Oak Ridge National Laboratory,*
Oak Ridge, Tennessee 37831, USA

(Received 7 November 2025; accepted 17 February 2026; published 1 April 2026)

Data-driven approaches using deep learning are emerging as powerful techniques to extract non-Gaussian information from cosmological large-scale structure. This work presents the first simulation-based inference (SBI) pipeline that combines weak lensing and galaxy clustering maps in a realistic Dark Energy Survey Year 3 (DES Y3) configuration and serves as preparation for a forthcoming analysis of the survey data. We develop a scalable forward model based on the COSMOGRIDV1 suite of N-body simulations to generate over one million self-consistent mock realizations of DES Y3 at the map level. Leveraging this large dataset, we train deep graph convolutional neural networks on the full survey footprint in spherical geometry to learn low-dimensional features that approximately maximize mutual information with target parameters. These learned compressions enable neural density estimation of the implicit likelihood via normalizing flows in a ten-dimensional parameter space spanning cosmological wCDM, intrinsic alignment, and linear galaxy bias parameters, while marginalizing over baryonic, photometric redshift, and shear bias nuisances. To ensure robustness, we extensively validate our inference pipeline using synthetic observations derived from both systematic contaminations in our forward model and independent BUZZARD galaxy catalogs. Our forecasts yield significant improvements in cosmological parameter constraints, achieving 2–3× higher figures of merit in the Ω_m-S_8 plane relative to our implementation of baseline two-point statistics and effectively breaking parameter degeneracies through probe combination. These results demonstrate the potential of SBI analyses powered by deep learning for upcoming stage-IV wide-field imaging surveys.

DOI: 10.1103/3sj1-1l9f

I. INTRODUCTION

The large-scale structure (LSS) of the mass distribution in the Universe encodes a wealth of information about its late-time evolution, offering a unique test bed for cosmological theories. Two primary observational probes of the LSS are "weak gravitational lensing" and "galaxy clustering." Weak lensing directly measures the projected matter density through coherent distortions of galaxy shapes by intermediate mass, while galaxy clustering tracks the distribution of luminous matter, which serves as a biased tracer of the underlying dark matter density. These probes of the LSS constrain cosmological parameters including the present-day matter density fraction Ω_m, the variance of linear density perturbations on the scale of 8 Mpc/h denoted σ_8, and the dark energy equation-of-state parameter w (see e.g. [1–4] for reviews). In addition, these observations are sensitive to astrophysical model parameters including intrinsic alignments and galaxy biasing. Crucially, combined analyses of weak lensing and galaxy clustering break parameter degeneracies that would otherwise limit the constraining power of each probe individually [5–8].

In recent years, observing programs targeting the LSS like the Dark Energy Survey (DES) [9], the Kilo-Degree Survey (KiDS) [10], the Subaru Hyper Suprime-Cam [11], and the Decam Local Volume Exploration Survey [12] have measured hundreds of millions of galaxy positions and shapes over thousands of square degrees of the sky, enabling cosmological parameter measurements with sub-5% precision [5,6,13,14]. Successor stage-IV surveys like

---

[*] Contact author: athomsen@phys.ethz.ch

the Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST) [15], the Euclid mission [16], and the Roman Space Telescope [17] are set to measure orders of magnitude larger datasets and expected to further increase the measurement precision to the subpercent level.

At sufficiently large scales, the projected matter density is well described as an isotropic Gaussian random field (GRF) [18]. Under this assumption, two-point statistics become sufficient summaries that capture all available statistical information [19]. This has contributed to their widespread use in cosmological analyses, for example, implemented as the correlation function in real space (e.g. [20,21]) or the power spectrum in harmonic space (e.g. [21,22]).

However, due to nonlinear structure formation and baryonic effects, the fields considered in this work only resemble GRFs on the largest scales and contain "non-Gaussian" features at intermediate and small scales [23]. Two-point statistics cannot capture this information, rendering them statistically insufficient. A growing body of literature therefore explores alternative summary statistics designed to extract the non-Gaussian component (see e.g. [24] for an overview in weak lensing) from the "map level." Within DES, examples of weak lensing analyses employing such summary statistics include peak counts [25–27], higher-order moments of weak lensing mass maps [28–31], higher-order correlation functions [32–34], the cumulative distribution function [35], wavelet harmonics [30,31], scattering transforms [30,31], and persistent homology [36]. Rather than predefining the summary statistic, an alternative approach parametrizes it as a (typically convolutional) neural network that is trained to automatically extract informative features [8,27,37–48]. As these methods mature, several have progressed beyond being proofs of concept to yield cosmological constraints from the Dark Energy Survey's first three years (DES Y3) of weak lensing observations [26,27,29,31,34,36].

A key challenge in utilizing these statistics is the general lack of analytical predictions relating them to the cosmological parameters. Moreover, the functional form of the likelihood is typically unknown. Simulation-based inference (SBI) addresses both issues by using large ensembles of simulations across different parameter values to establish the parameter-statistic relationship [49]. Neural density estimation [50,51] can then be used to learn the probability density of interest, in our case the likelihood, directly from the simulated data [52–54].

We present an SBI pipeline that employs map-level neural network summary statistics to constrain cosmology by jointly conditioning on DES Y3 weak lensing and galaxy clustering data. This represents the first application of higher-order summary statistics to this probe combination within the DES. In this paper, we validate our methodology on simulations, with the application to the actual DES Y3 observation to be presented in a forthcoming companion paper.

This paper is organized as follows. Section II provides a brief overview of the blinded DES Y3 source and lens galaxy catalogs. Section III describes our forward model, which transforms the dark matter particle light cones from the COSMOGRIDV1 [23] simulation suite into more than one million self-consistent DES Y3 weak lensing and galaxy clustering mocks varying w cold dark matter (wCDM), baryonification, intrinsic alignment, linear galaxy biasing, and observational nuisance parameters. The SBI methodology is detailed in Sec. IV, including our map-level compression networks, the two-point statistic baseline, and neural density estimation using normalizing flows. Section V introduces the BUZZARD [55] mock catalogs, which are external to our forward model and serve as an independent validation dataset. Section VI explains how we use these and other mocks to test our pipeline's robustness against model misspecification. We present our results in Sec. VII, contrasting our map-level compression with the two-point baseline, as well as comparing the different cosmological probes individually and in combination. Finally, we summarize our findings and conclusions in Sec. VIII and discuss future prospects for this methodology.

II. SURVEY CONFIGURATION

A. Dark Energy Survey Year 3

The DES [56,57] is an observational program that imaged roughly 5000 deg^2 of the southern hemisphere over six years (2013–2019) providing photometric measurements in the five optical near-infrared broadbands *grizY*. The measurements were taken with the 570 megapixel dark energy camera [58] mounted on the 4 m Blanco telescope at Cerro Tololo Inter-American Observatory in Chile. For details on the image processing pipeline, we refer the interested reader to [59].

We utilize the DES Y3 data from the first three years of operations, which have been made publicly available as the DES Data Release 1 [60]. However, in this work, we only employ selected properties of the observed DES Y3 catalogs such as the masking, galaxy number densities, redshift distributions, and calibration properties to conduct a realistic forecast and validate the end-to-end inference pipeline against various synthetic mocks; no cosmological constraints are derived from the real observations here. We leave these cosmological results to a later companion paper.

B. Source galaxy catalog

The DES Y3 weak lensing shape catalog [61] contains approximately 100×10^6 source galaxies within an area of 4143 deg^2, yielding a weighted number density of $n_{\mathrm{eff}} = 5.59\ \mathrm{gal/arc\,min}^2$.

For the catalog, the galactic ellipticities or shears are measured from the observed (multiband) noisy images using the self-calibrating Metacalibration algorithm [62,63]. Additional redshift-dependent detection and blending

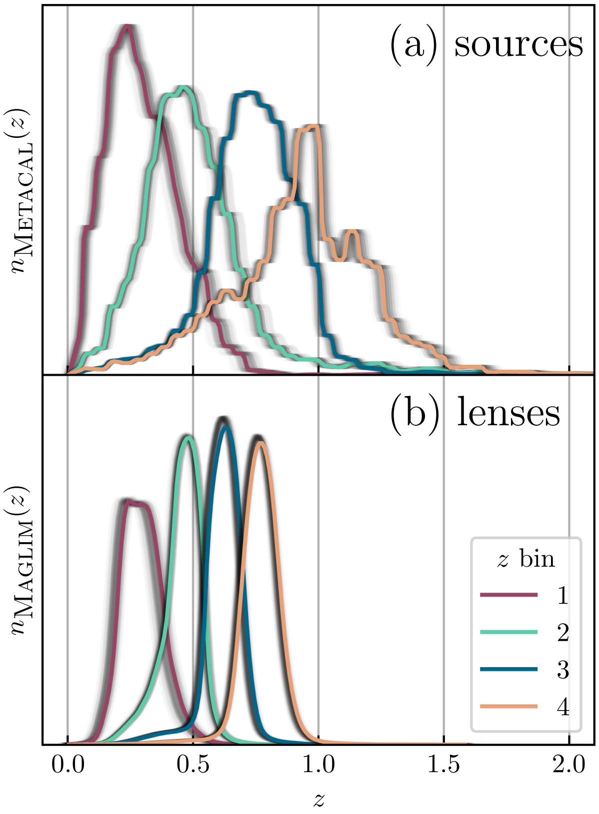

FIG. 1. Normalized redshift distributions of (a) the Metacalibration source galaxy sample used for weak lensing and (b) the MAGLIM lens galaxy sample used for galaxy clustering. The colored lines indicate the base distributions, while the partially transparent overlapping gray lines illustrate the redshift uncertainty via 50 draws from the photo-z distributions parametrized in Eqs. (5) and (6) according to Table I.

effects identified in [64] through end-to-end image simulations, which are not accounted for by Metacalibration, can be modeled as a multiplicative bias at the 2%–3% level; see Sec. III D 4 for our treatment.

The photometric redshifts of the objects in the catalog are determined by the SOMPZ algorithm [65] and further calibrated in [64,66]. The galaxies are subdivided into four tomographic bins of roughly equal number density [65], resulting in redshift distributions we denote as $n^i_{\text{METACAL}}(z)$ for $i \in \{1,2,3,4\}$ and plot in Fig. 1(a).

For this paper, the catalog is utilized only for its mask, redshift distributions, and in the shape noise generation, where randomly sampled and rotated galaxies from the catalog are used to generate noise maps as described in Sec. III D 2; the original shape catalog is not used.

C. Lens galaxy catalog

For the lens galaxies, we choose to use the DES Y3 magnitude limited sample [67] denoted as MAGLIM and employed as fiducial for the DES Y3 3×2 pt analysis [68]. The sample is characterized by an upper magnitude cut in the i-band that depends linearly on the photometric redshift estimate z from the directional neighborhood fitting (DNF) algorithm [69,70] like $i < 4z + 18$, which is a selection that has been optimized for wCDM constraints by balancing the resulting galaxy number density and accuracy of the photometric redshift estimates [67]. The catalog contains around 10.7×10^6 galaxies grouped into six tomographic bins with redshift distributions $n^i_{\text{MAGLIM}}(z)$ [see Fig. 1(b)], of which we discard the last two following the fiducial analysis in [5].

Just as with the source galaxy sample, we do not use the original lens galaxy catalog in this work; for the purposes of forecasting and validation, we only consider the catalog's mask, mean galaxy number density, and redshift distributions.

III. FORWARD MODELING

In this section, we present our forward model for self-consistent simulated DES Y3-like "weak lensing" and "galaxy clustering" maps. These maps respectively resemble the source and lens galaxy samples described in Sec. II, matching their masking, average number density, and redshift distributions while varying cosmological wCDM parameters, baryonic feedback effects, intrinsic alignment contributions, linear galaxy bias, and observational nuisances.

This forward modeling approach offers several advantages. First, observational systematics such as masking, shear biases, and redshift errors can be directly incorporated despite being difficult to treat analytically even at the two-point level [71–73]. Second, within the simulation-based inference framework, map-level forward modeling enables training neural compression networks that potentially capture the full information content of the pixelized fields without relying on handcrafted summary statistics sensitive only to specific features of the data. The compression networks thus serve as non-Gaussian summary statistics for which no analytical connection to cosmological theory exists; instead, the connection is established through numerical forward simulations.

Both training of these compression networks and inference require large numbers of such simulations, motivating our use of the COSMOGRIDV1 suite ([23], hereafter K23), publicly available via [74]. A schematic overview of the key postprocessing steps applied to these simulations is provided in Fig. 2, with the following subsections following this data flow.

The implementation of the forward model is available on GitHub [75].

A. COSMOGRIDV1 simulations

The forward model is based on the COSMOGRIDV1 simulation suite of flat wCDM cosmologies. The dark-matter-only N-body simulations were run with the

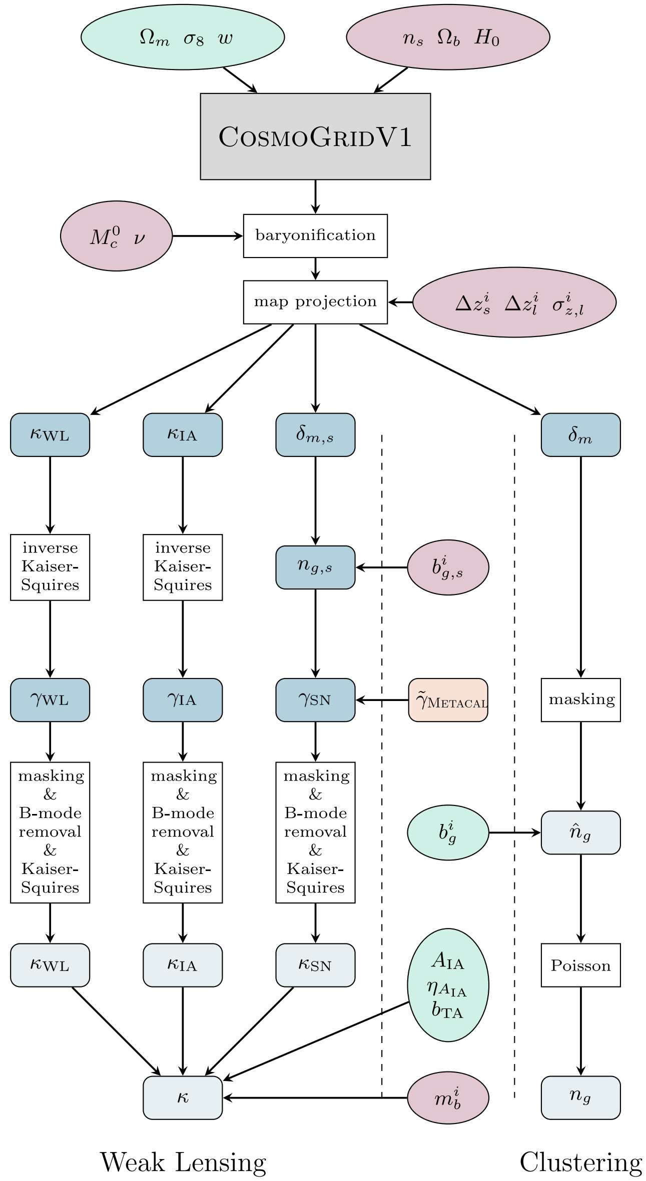

FIG. 2. Schematic overview of the processing pipeline we apply to transform (baryonified) particle shells from the COSMOGRIDV1 simulations into mock weak lensing and galaxy clustering maps matching selected DES Y3 properties. Sharp-cornered boxes represent processing steps, while blue rounded boxes denote (full- or partial-sky) HEALPIX maps. Green and red ellipses indicate constrained and marginalized parameters, respectively. The tomographic bin index $i \in \{1, 2, 3, 4\}$ for source (subscript s) and lens (subscript l) galaxy samples is omitted from map names for simplicity. The abbreviation "WL" indicates weak lensing signal, "IA" intrinsic alignment, and "SN" shape noise. We denote the scrambled shear catalog of randomly rotated source galaxies used to generate shape noise maps as $\tilde{\gamma}_{\rm METACAL}$ and omit the $\kappa_{\rm IA,TA}$ [see Eq. (7)] map for clarity.

open-source PKDGRAV3 code [76], which is optimized for hybrid CPU-GPU cluster architectures and has computational cost that scales linearly with the number of particles.

The following subsections describe the suite in more detail.

1. Cosmological parameters

The simulations comprising the COSMOGRIDV1 suite sample the parameter space of *w*CDM universes. A flat universal geometry is maintained throughout by setting the dark energy density parameter to $\Omega_\Lambda = 1 - \Omega_m$. The neutrino mass is fixed to three degenerate species with $m_\nu = 0.02$ eV each, giving a total mass sum of $\sum m_\nu = 0.06$ eV.

While recent results, including those from DES [77] and the Dark Energy Spectroscopic Instrument (DESI) [78–80], suggest possible time variation in the dark energy equation of state (parametrized by w_0 and w_a), our analysis employs the *w*CDM model with constant w, as the COSMOGRIDV1 was designed around this choice. We defer exploration of dynamical dark energy to future work, which will require generating a new suite of simulations.

The COSMOGRIDV1 contains two main sets of simulations, defined by the sampled point(s) in parameter space, which we generally denote as θ:

(a) Fiducial: The fiducial cosmological parameters of the COSMOGRIDV1 are based on the Planck2018 results [81] and are shown in the $\theta_{\rm fid}$ column of Table I. There is a total of 200 independent simulation runs at the fiducial cosmology, from which we build 1000 semi-independent realizations as described in Sec. III B 1.

Note that in a traditional likelihood analysis, realizations at the fiducial cosmology are typically used to estimate the summary statistic's covariance matrix. However, this is not necessary in the simulation-based inference framework we employ in this work, as an approximation to the likelihood function is learned directly by a normalizing flow (Sec. IV B).

(b) Grid: In simulation-based inference, the grid in parameter space establishes the necessary connection between the cosmological and astrophysical variables making up the vector θ to be inferred and the (potentially synthetic) observable.

The parameter grid within the COSMOGRIDV1 follows a Sobol sequence [82], which is a quasirandom, low-discrepancy sampling scheme that fills space uniformly. The dimensionality of the sequence is expandable beyond the six cosmological parameters initially included in the COSMOGRIDV1: Ω_m (matter density), σ_8 (matter clustering amplitude), w (dark energy equation of state), n_s (spectral index), Ω_b (baryon density), and H_0 (Hubble constant). We use this property to extend the sequence to include baryonification parameters (Sec. III A 3).

All one-dimensional priors over these parameters are uniform and flat, forming a hyperrectangle in aggregate, except for additional restrictions applied in the Ω_m-σ_8 (Fig. 3) and Ω_m-w (Appendix A of [40], hereafter F22) planes. The 2500 grid points are evenly divided between

TABLE I. Cosmological, astrophysical, and nuisance parameters of the fiducial (θ_{fid}) and grid (wide and narrow prior) COSMOGRIDV1 subsets. The elements of a joint, multivariate Sobol sequence are represented as $\mathcal{S}[a, b]$, where the square brackets define the bounding interval. Similarly, joint latin hypercube sampling is denoted by $\mathcal{L}[a, b]$. Univariate normal distributions with mean μ and standard deviation σ are designated as $\mathcal{N}(\mu, \sigma)$. We marginalize the cosmological parameters n_s, Ω_b, H_0, the baryonification variables, and photometric redshift and shear bias nuisances.

θ	θ_{fid}	Wide prior	Narrow prior
Cosmology			
Ω_m	0.26	$\mathcal{S}[0.1, 0.5]$	$\mathcal{S}[0.15, 0.45]$
σ_8	0.84	$\mathcal{S}[0.4, 1.4]$	$\mathcal{S}[0.5, 1.3]$
w	-1	$\mathcal{S}[-2, -0.33]$	$\mathcal{S}[-1.25, -0.75]$
n_s	0.9649	$\mathcal{S}[0.87, 1.07]$	$\mathcal{S}[0.93, 1]$
Ω_b	0.0493	$\mathcal{S}[0.03, 0.06]$	$\mathcal{S}[0.04, 0.05]$
H_0	67.3	$\mathcal{S}[64, 82]$	$\mathcal{S}[65, 75]$
Baryonification			
$\log(M_c^0)$	13.82	$\mathcal{S}[12, 15]$	…
ν	0	$\mathcal{S}[-2, 2]$	…
Intrinsic alignment			
A_{IA}	0.5	$\mathcal{L}[-3, 3]$	…
$\eta_{A_{\text{IA}}}$	1.5	$\mathcal{L}[-4, 6]$	…
b_{TA}	1	$\mathcal{L}[0, 2]$	…
Galaxy biasing			
b_g^1	1.34	$\mathcal{L}[0.8, 3]$	…
b_g^2	1.42	$\mathcal{L}[0.8, 3]$	…
b_g^3	1.50	$\mathcal{L}[0.8, 3]$	…
b_g^4	1.57	$\mathcal{L}[0.8, 3]$	…
Source photo-z			
$\Delta z_s^1 \times 10^2$	…	$\mathcal{N}(0, 1.8)$	…
$\Delta z_s^2 \times 10^2$	…	$\mathcal{N}(0, 1.5)$	…
$\Delta z_s^3 \times 10^2$	…	$\mathcal{N}(0, 1.1)$	…
$\Delta z_s^4 \times 10^2$	…	$\mathcal{N}(0, 1.7)$	…
Source shear bias			
$m_b^1 \times 10^2$	…	$\mathcal{N}(-0.6, 0.9)$	…
$m_b^2 \times 10^2$	…	$\mathcal{N}(-2.0, 0.8)$	…
$m_b^3 \times 10^2$	…	$\mathcal{N}(-2.4, 0.8)$	…
$m_b^4 \times 10^2$	…	$\mathcal{N}(-3.7, 0.8)$	…
Lens photo-z			
$\Delta z_l^1 \times 10^2$	…	$\mathcal{N}(-0.9, 0.7)$	…
$\Delta z_l^2 \times 10^2$	…	$\mathcal{N}(-3.5, 1.1)$	…
$\Delta z_l^3 \times 10^2$	…	$\mathcal{N}(-0.5, 0.6)$	…
$\Delta a_l^4 \times 10^2$	…	$\mathcal{N}(-0.7, 0.6)$	…
$\sigma_{z,l}^1$	…	$\mathcal{N}(0.98, 0.06)$	…
$\sigma_{z,l}^2$	…	$\mathcal{N}(1.31, 0.09)$	…
$\sigma_{z,l}^3$	…	$\mathcal{N}(0.87, 0.05)$	…
$\sigma_{z,l}^4$	…	$\mathcal{N}(0.92, 0.05)$	…

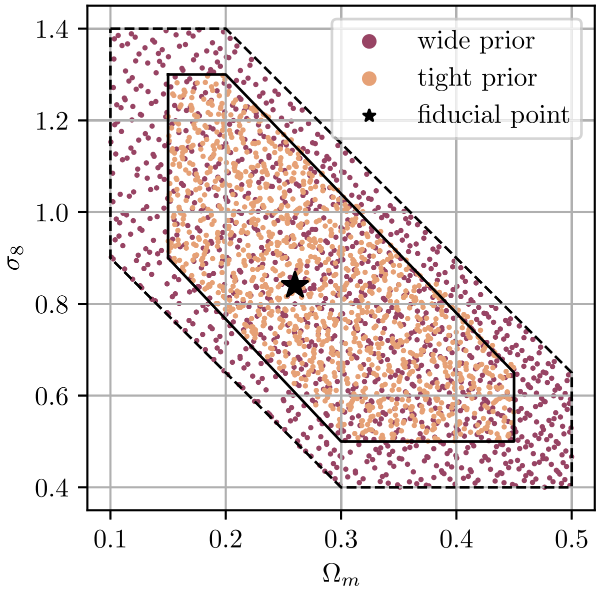

FIG. 3. Projection of the fiducial and 2500 grid cosmologies included in the COSMOGRIDV1 to the Ω_m-σ_8 plane. Dashed and solid lines represent the wide and narrow priors, respectively, with the corresponding points colored in blue and orange. The black star marks the fiducial cosmology.

narrow and wide priors with 1250 cosmologies each, where the narrow prior provides higher sampling density in the region of greatest interest. For the forecasts conducted in this work, we consider the entire parameter range of the wide prior and the narrow prior therefore has no further Bayesian significance. The wide and narrow prior parameter ranges are listed in Table I.

For every grid point, the COSMOGRIDV1 includes seven independent simulation runs stemming from different random initial conditions. In the later map projection step, we mix these seven runs to yield 20 permutations as outlined in Sec. III B 1.

2. *Configuration*

Throughout the COSMOGRIDV1, simulation outputs are stored in light cone format as thin shells of pixelwise dark matter particle counts in HEALPIX [83] maps (detailed in [84]) with resolution $n_{\text{side}} = 2048$, corresponding to an angular scale of approximately 1.72 arc min.

Table II summarizes the main simulation settings. A distinction is made between:

(a) Main settings: The main simulation settings apply to the "fiducial" and "grid" subsets. The configuration was informed by trade-offs between accuracy and computational feasibility, aiming to facilitate analyses of stage-III LSS surveys like the one conducted in this work (K23).

(b) Benchmark settings: The COSMOGRIDV1 contains a set of benchmark simulations designed to test and evaluate the robustness of the analysis with respect to the main simulation settings. These include (i) larger box sizes at equal particle density to verify the adequacy of the box replication scheme, (ii) greater numbers of

TABLE II. Summary of the main simulation settings for different subsets of the COSMOGRIDV1. The number of distinct cosmologies (or sampled points in parameter space) is denoted by N_{cosmos}, the number of permutations per cosmology by N_{perts}, the nonreplicated simulated box size by L_{box}, the number of simulated particles by N_{part}, and the number of redshift shells (or stored time steps) by $N_{z\text{-shells}}$.

	N_{cosmos}	N_{perts}	L_{box} (Mpc/h)	N_{part}	$N_{z\text{-shells}}$
Main					
Fiducial	1	1000	900	832^3	69
Grid	2500	20	900	832^3	69
Benchmark					
Base	1	20	900	832^3	69
L_{box}	1	20	2250	2080^3	69
N_{part}	1	20	900	2048^3	69
No. z-shells	1	20	900	832^3	392

particles to decrease the amount of shot noise, and (iii) increased numbers of stored redshift shells in the light cone improving the resolution in the radial direction. All benchmark simulations were conducted at the fiducial cosmological parameters and share initial conditions with runs performed using the main simulation settings in order to fix cosmic variance in direct comparisons.

3. Baryonic feedback

The dark matter particles in PKDGRAV3 evolve under gravity alone. However, it has been shown that baryonic feedback can bias cosmological constraints, particularly on small scales (e.g. [85–88]).

The standard DES Y3 modeling strategy [7] addresses this through scale cuts, removing small scales affected by baryonic physics as modeled in the active galactic nuclei simulation [85] from the hydrodynamic Overwhelmingly Large Simulations suite [89]. These scale cuts are tuned to ensure posterior shifts with respect to a fiducial uncontaminated data vector remain below 0.3σ in the Ω_m-S_8 plane.

In this work, we instead incorporate baryonic effects directly into our modeling by postprocessing the COSMOGRIDV1 light cone shells with the effective "baryonification model" developed in [90–92], which displaces the dark matter particles according to a physically motivated prescription. The interested reader is directed to [91] for a comprehensive explanation of the original model and to K23 for details on the slightly modified, shell-level implementation used to baryonify the COSMOGRIDV1.

Following F22, we only vary the model parameter M_c defining the mass dependence of the gas profile, as it has been shown to have the biggest impact on cosmology [93], and assume a power-law redshift dependence

$$M_c = M_c^0 (1+z)^{\nu}$$

in terms of the new model parameters M_c^0 and ν. We assign every grid cosmology unique values M_c^0 and ν by extending the Sobol sequence with two additional dimensions, which we scale according to the prior ranges given in Table I. Since the COSMOGRIDV1 only resolves halos with masses down to approximately $10^{13} M_{\odot}/h$, values at the lower end of the prior interval result in negligible baryonification. All other parameters in the baryonification model are fixed to the same values as in F22, which are motivated by x-ray observations and listed in Table 2 of [91].

Throughout this work, we use the baryonified COSMOGRIDV1 particle shells unless stated otherwise and marginalize over M_c^0 and ν. We assess the impact of baryonic effects at our fiducial scale cuts in Sec. VI A 2.

B. Map projection

We project the particle shells of the simulated light cones onto probe maps with the publicly available [94] UFALCON code [95–97], which has been used in several forecasts [8,37,97–99] and cosmological inferences from real observational data [26,38,40,100,101]. The code employs the Born approximation, which was also assumed in [27,30,31] and shown to be sufficiently precise for stage-III surveys like DES in [102] and the systematics testing in [38], an application to KiDS-450 weak lensing data.

As described in more detail in K23, each pixel in the projected HEALPIX probe map m is computed as

$$m^{\mathrm{pix}} \approx \sum_b W_m \int_{\Delta z_b} \frac{\mathrm{d}z}{E(z)} \delta_{3\mathrm{D}} \left[\frac{c}{H_0} \mathcal{D}(z) \hat{n}^{\mathrm{pix}}, z \right], \tag{1}$$

where the index b runs over the light cone's redshift shells of thickness Δz_b, W_m is the kernel associated with a given probe as defined in the following, $\delta_{3\mathrm{D}}$ is a Dirac delta function, c is the speed of light, $\mathcal{D}(z)$ is the dimensionless comoving distance, $E(z) \coloneqq \mathrm{d}z/\mathrm{d}\mathcal{D}$ is the dimensionless Hubble parameter, and $\hat{n}^{\mathrm{pix}}$ is a unit vector pointing to the pixel's center.

Following [8,38,40], we define the probe kernels of the weak lensing (first introduced in [103]) and intrinsic alignment (see Sec. III D 3) signal as

$$W_{\mathrm{WL}} = \frac{3}{2} \Omega_{\mathrm{m}} \frac{\int_{\Delta z_b} \frac{\mathrm{d}z}{E(z)} \int_z^{z_s} \mathrm{d}z' n(z') \frac{\mathcal{D}(z)\mathcal{D}(z,z')}{\mathcal{D}(z')} \frac{1}{a(z)}}{\int_{\Delta z_b} \frac{\mathrm{d}z}{E(z)} \int_{z_0}^{z_s} \mathrm{d}z' n(z')}, \tag{2}$$

$$W_{\mathrm{IA}} = - \frac{\int_{\Delta z_b} \mathrm{d}z C_1 \rho_{\mathrm{crit}} \frac{\Omega_m}{D_+(z)} n(z)}{\int_{\Delta z_b} \frac{\mathrm{d}z}{E(z)} \int_{z_0}^{z_s} \mathrm{d}z' n(z')}, \tag{3}$$

where $n(z)$ is a normalized redshift distribution, z_s and z_0 denote the source and observer redshifts, respectively, $a(z)$ is the scale factor, $C_1 = 5 \times 10^{-14} h^{-2} M_{\odot}\ \mathrm{Mpc}^3$ is a normalization constant [104], $\rho_{\rm crit}$ is the critical density today, and $D_+(z)$ is the normalized linear growth factor with $D_+(0) = 1$. We evaluate Eq. (1) for these kernels and $n_{\rm METACAL}(z)$, resulting in the full-sky convergence maps $m = \kappa_{\rm WL}, \kappa_{\rm IA}$ in the left part of Fig. 2.

The kernel used to project the full-sky linear matter density contrast maps is defined as

$$W_{\rm GC} = \frac{\int_{\Delta z_b} {\rm d}z n(z)}{\int_{\Delta z_b} \frac{{\rm d}z}{E(z)} \int_{z_0}^{z_s} {\rm d}z' n(z')}. \tag{4}$$

Insertion into Eq. (1) yields the map $m = \delta_m$ for $n_{\rm MAGLIM}(z)$ (used for lens galaxy clustering maps; Fig. 2 right, Sec. III E) and $m = \delta_{m,s}$ for $n_{\rm METACAL}(z)$ (used for shape noise generation and modeling of intrinsic alignment of the source galaxy sample; Fig. 2 left, Secs. III D 2 and III D 3).

During the projection step, we downsample the map resolution to HEALPIX $n_{\rm side} = 512$ (corresponding to an angular pixel size of approximately 6.87 arc min), which reduces the, nevertheless substantial, storage and compute requirements of this work. The decreased resolution also acts as a low-pass filter, removing small scales. However, due to limitations of our physics modeling validated in Sec. VI A, we apply additional scale cuts to erase further small-scale information as detailed in Sec. III F.

1. Shell permutations

To make the N-body simulations computationally feasible for the large cosmological volumes necessary at high redshifts, PKDGRAV3 implements a box replication scheme. This can introduce unwanted artifacts like discontinuities (K23) and underestimation of cosmic variance on the largest scales [38]. To avoid the former effect, we apply the "shell permutation scheme" introduced in K23 during the map projection step.

As an additional benefit, this procedure increases the number of available pseudoindependent realizations as for a fixed cosmology at a time, simulation boxes stemming from different independent runs are randomly combined. We refer the interested reader to Sec. IV. 1 in K23 for further details.

In this work, the seven independent runs of the "grid" subset of the COSMOGRIDV1 are mixed to yield 20 permutations per cosmology, while the 200 simulation runs at the fiducial cosmology are combined to 1000 distinct permutations of that ordered index set; see the $N_{\rm perts}$ column in Table II.

2. Redshift errors

Following the analysis in [5], we model the uncertainty associated with the redshift distribution of tomographic bin i for the source and lens galaxy samples using the "shift" parameters Δz_s^i and Δz_l^i, respectively, and the "stretch" parameter $\sigma_{z,l}^i$ for the lenses only. The parameter distributions are assumed to be Gaussian with mean and standard deviation as reported in the "photo-z" rows of Table I.

We do not aim to constrain the shift and stretch uncertainty parameters; instead, we treat them as nuisances to be marginalized. To this end, we employ the same general strategy as F22 and [27]: First, we draw samples from the prescribed Gaussian distributions. Then, we project the probe maps using the altered redshift distributions

$$n_{\rm METACAL}^i(z) = \hat{n}_{\rm METACAL}^i(z - \Delta z_s^i) \tag{5}$$

for the source sample and

$$n_{\rm MAGLIM}^i(z) = \frac{1}{\sigma_{z,l}^i} \hat{n}_{\rm MAGLIM}^i \left(\frac{z - \langle z \rangle - \Delta z_l^i}{\sigma_{z,l}^i} + \langle z \rangle \right) \tag{6}$$

for the lens sample, where $\hat{n}$ denotes the original, noiseless distribution and $\langle z \rangle$ its mean redshift [68]. Thus, the redshift uncertainty directly enters the maps we create.

It can be shown [105] that this implicit marginalization is mathematically equivalent to the standard integral approach.

3. Validation

The power spectra of the noiseless full-sky probe maps from the COSMOGRIDV1 have been validated against theoretical predictions from two independent codes. In F22, comparison was performed with the `PyCCL` package [106] described in [107] for Eqs. (2) and (3) using KiDS-1000 redshift distributions. In K23, validation was conducted against the PYCOSMO package [108] introduced in [109] for Eqs. (2)–(4) using generic stage-III-like redshift distributions. Both studies showed agreement within 5%, consistent with the minimal error expected from discrepancies between different theory predictions [110,111].

C. Masking and padding

We apply realistic masking to our forward-modeled maps to ensure they resemble the DES Y3 observations, which are contained within a complex footprint that excludes numerous intermediate objects, such as stars in the Milky Way.

To be able to cut out a total of four independent (except for supersurvey modes, which we handle as described in Sec. III F 2) DES Y3 footprints from our full-sky simulations, we first rotate the fiducial mask from [68] by $-7.16°$ along the y-axis and $-69.9°$ along the z-axis as illustrated in Fig. 4(a). At $n_{\rm side} = 512$, this yields a mask of 4178 deg^2.

From this modified orientation, we apply HEALPIX octahedral symmetry transforms (rotations by 90° or mirroring along the equator) to the footprint to obtain

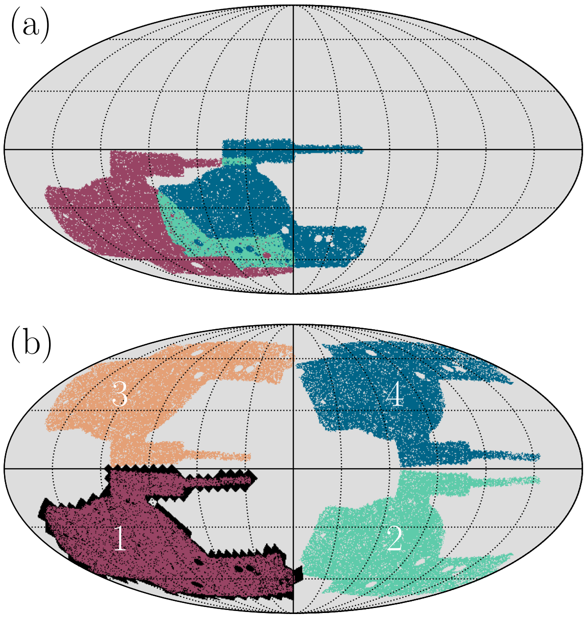

FIG. 4. Full-sky Mollweide projection of (a) how the original DES Y3 footprint (blue) is rotated to a position (red) that allows for (b) four nonoverlapping cutouts. These distinctly colored patches are related by HEALPIX symmetries such that there is a perfect one-to-one correspondence between the pixels. The black padding with zeros along patch 1 is determined by the lowest $n_{\rm side}$ used within the DEEPSPHERE networks; $n_{\rm side} = 16$ here. It is not part of the original survey area.

three more nonoverlapping positions of the DES mask on the celestial sphere. These transforms maintain a perfect bijection between pixels, preventing the introduction of unwanted artifacts. Thereby, the four footprints are fully equivalent. Figure 4(b) illustrates these additional orientations and the black padding applied to the base footprint. This zero padding is required to ensure that the DEEPSPHERE networks [112,113] described in Sec. IV A 2 can internally downsample to a minimal HEALPIX $n_{\rm side}$ resolution without error.

As illustrated in Fig. 2 and detailed in the following Sec. III D, we deliberately apply the masking to the physical fields that are directly accessible observationally (i.e. shear for the weak lensing maps) to ensure that the forward model remains consistent between our COSMOGRIDV1 mock maps and (synthetic) observations constructed from a galaxy catalog.

D. Weak gravitational lensing

Weak lensing measurements enable the reconstruction of the convergence field, which represents a weighted projection of the intervening matter density distribution between the observer and background source galaxies along the line of sight. Since these maps directly trace the underlying matter distribution (predominantly dark matter), they are commonly referred to as mass maps [114].

We forward-model DES Y3-like convergence maps κ by summing contributions from the noiseless weak lensing signal, intrinsic alignment, and shape noise,

$$\kappa^i = \kappa^i_{\rm WL} + A^i_{\rm IA}(1 + b_{\rm TA}\delta_{m,s})\kappa^i_{\rm IA} + \kappa^i_{\rm SN}, \tag{7}$$

where i indexes the redshift bin, $A_{\rm IA}$ and $b_{\rm TA}$ are free parameters of the intrinsic alignment model, and $\delta_{m,s}$ is the matter density contrast map for the source galaxy redshift distribution. Each field component is detailed in the following subsections.

1. Mass mapping and B-mode removal

The convergence field κ is not directly observable but can be reconstructed from the noisy shear field γ obtained from source galaxy ellipticity measurements. To ensure consistent processing (see below), we forward model our weak lensing mass maps through a series of conversions between convergence and shear ($\kappa \to \gamma \to \kappa$), as illustrated in the left side of Fig. 2.

In the first step of the forward modeling pipeline, we project full-sky convergence maps $\kappa_{\rm WL}$ and $\kappa_{\rm IA}$ from simulated dark matter particle shells as described in Sec. III B. We then perform an inverse Kaiser-Squires transform [115,116] to convert these to shear maps $\gamma_{\rm WL}$ and $\gamma_{\rm IA}$, respectively. This conversion to shear allows us to apply identical subsequent processing steps, including masking, to both the observed $\gamma_{\rm obs}$ and forward-modeled maps.

By construction, the $\gamma_{\rm WL}$ and $\gamma_{\rm IA}$ obtained from simulated κ-maps via inverse Kaiser-Squires transform contain only E-modes, unlike $\gamma_{\rm obs}$ and the shape noise map $\gamma_{\rm SN}$ described in the following section, which may contain B-modes. A B-mode mismatch between forward-modeled and observed shear maps can introduce biases in the inference. Following the "mode removal" procedure in F22 and [38], we hence remove all potential B-modes from the γ-maps after applying masking. This ensures that forward-modeled mocks, noise realizations, and real observations all contain the same masking effects and only E-modes.

Lastly, to reduce storage requirements by half, we apply a direct Kaiser-Squires transform to the masked spin-2γ-maps to convert back to scalar κ-maps. The spherical harmonics decomposition involved in this transform is only defined on the full sky, which we address by padding the area outside the survey footprint with zeros.

2. Shape noise

We generate shape noise stemming from the unknown intrinsic ellipticities of the source galaxies self-consistently using the dark matter density contrast maps $\delta_{m,s}$ and the shuffled shear catalog $\tilde{\gamma}_{\rm METACAL}$ (with positions discarded) of the source galaxies through the following steps:

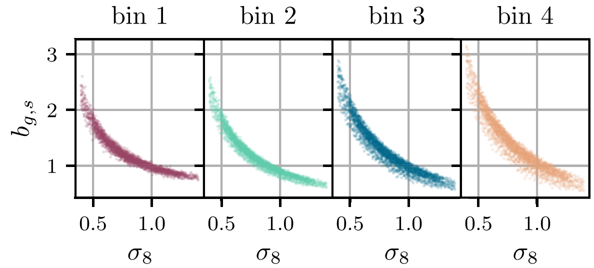

FIG. 5. Scatter plot depicting the source galaxy bias $b_{g,s}$ we find for the 2500 unique cosmologies of the COSMOGRIDV1 by fitting the source galaxy number count histogram to a reference BUZZARD simulation of fixed cosmology.

First, we build the galaxy count map of tomographic bin i assuming a linear galaxy biasing model

$$n^i_{g,s} = \langle n^i_{g,s} \rangle \left(1 + b^i_{g,s} \frac{\delta^i_{m,s} - \langle \delta^i_{m,s} \rangle}{\langle \delta^i_{m,s} \rangle}\right),$$

where $\langle n_{g,s} \rangle$ is the mean number of source galaxies per pixel and $b_{g,s}$ is the linear galaxy clustering bias associated with the source galaxy sample. We determine this bias by matching the pixel histogram of galaxy counts for each simulated cosmology to a reference BUZZARD simulation (Sec. V B) serving as our synthetic mock observation. The resulting dependence is shown in Fig. 5. By varying $b_{g,s}$ per cosmology in this way, the cosmology dependence of the one-point source galaxy distribution is absorbed into the bias parameter, effectively suppressing it in the resulting source galaxy maps. The source galaxy bias is therefore separately fit to the (synthetic) observational data and fixed in the analysis rather than marginalized.

Given such a map of source galaxy counts, we generate a new catalog of pure shape noise by randomly sampling as many galaxies from the real DES Y3 source galaxy shear catalog $\tilde{\gamma}_{\text{METACAL}}$ as there are galaxies in any given pixel according to $n_{g,s}$. In this step, we discard any positional information from $\tilde{\gamma}_{\text{METACAL}}$ and apply random rotations to the galaxies, such that any weak lensing signal is scrambled. The Metacalibration inverse variance weights are kept. Lastly, we obtain shape noise shear maps γ_{SN} by computing the pixelwise weighted mean of the shear over all contained galaxies.

This approach ensures that each pixel's shape noise level is coupled to its source galaxy count; pixels with more galaxies experience lower shape noise. By linking shape noise to local galaxy densities, we account for the noise component of source galaxy clustering effects, which measurably impact certain map-based statistics in DES Y3 data [117]. We neglect the contribution of source clustering to the weak lensing signal itself, as this has been shown to be small compared to the noise component [117].

We choose to generate ten shape noise realizations per noiseless signal map at the fiducial cosmology and five realizations each for the cosmologies forming the parameter grid.

3. Intrinsic alignment

Intrinsic alignment refers to the nonrandom, correlated orientations of galaxies due to their mutual gravitational interactions or shared formation history, which can bias weak lensing measurements by introducing additional alignments unrelated to the lensing signal [118].

The kernel in Eq. (3) constitutes a map-level implementation of the nonlinear alignment (NLA) model [104] and has been validated against other theory predictions in [23,40]. The NLA model is a special case ($a_2 = b_{\text{TA}} = 0$) of the more general tidal alignment and tidal torquing (TATT) model [119] that has been used in the fiducial DES Y3 3×2 pt analysis [68].

Following [8] (hereafter K22), we modify the standard redshift dependence of the NLA model to the effective prescription

$$A^i_{\text{IA}} = A_{\text{IA}} \int_z dz n^i_{\text{METACAL}}(z) \left(\frac{1+z}{1+z_{\text{pivot}}}\right)^{\eta_{A_{\text{IA}}}}, \quad (8)$$

producing one amplitude per tomographic bin i, where $z_{\text{pivot}} = 0.62$ is a pivot redshift taken from [68] and we integrate over the source galaxy redshift distribution $n^i_{\text{METACAL}}(z)$. This formulation preserves the underlying power-law dependence of the NLA model, but yields a single value per bin rather than evaluating it at the finer redshift resolution of the individual light cone shells in Eq. (1). We validate this minor modification in Sec. VI A 2.

In addition, we include the b_{TA} term of the TATT model

$$\kappa_{\text{IA,TA}} = A_{\text{IA}} b_{\text{TA}} \delta_{m,s} \kappa_{\text{IA}},$$

which encodes the coupling strength with the local density field [120] in Eq. (7). Our intrinsic alignment model therefore represents an intermediate approach between the NLA and TATT models.

For each cosmology or, equivalently, point on the Sobol sequence, we jointly sample the free intrinsic alignment parameters A_{IA}, $\eta_{A_{\text{IA}}}$, and b_{TA} using Latin hypercube sampling (LHS) with prior intervals listed in the "intrinsic alignment" rows of Table I.

Currently, a map-level implementation of the remaining terms of the TATT model is an open research question [23] and therefore not included in this work. Moreover, our use of the NLA model is motivated by the fiducial DES Y3 3×2 pt analysis [68], which found consistent cosmological constraints between the TATT and NLA models (Appendix E.4 in [68]) for the unblinded DES Y3 data.

4. Shear bias

In addition to the multiplicative shear biases corrected by the self-calibrating Metacalibration algorithm [62,63], there are redshift-dependent detection and blending effects found in [64] that can be modeled as a multiplicative bias on the 2%–3% level. In line with [27], we account for this uncertainty by sampling the multiplicative bias factor m_b^i from the normal distributions referenced in Table I and modify the forward-modeled convergence map as

$$\kappa_{\mathrm{WL}}^i = (1 + m_b^i)\hat{\kappa}_{\mathrm{WL}}^i$$

for each tomographic bin i, where $\hat{\kappa}_{\mathrm{WL}}^i$ is the original map lacking the random correction. This way, the m_b^i are treated as nuisance parameters and marginalized like the redshift errors in Sec. III B 2.

Furthermore, null tests in [28] demonstrated the shear catalog's robustness against additive biases; therefore, we do not include them in this work.

E. Galaxy clustering

The galaxy clustering maps n_g count the number of lens galaxies in each pixel. To forward model such maps from the COSMOGRIDV1 simulations, we first compute dark matter density contrast maps δ_m by projecting the particle shells according to the kernel in Eq. (4). We then require a model to relate the matter and galaxy distributions, for which we adopt a linear biasing prescription as defined below.

The adequacy of this approach is validated in Sec. VI A for the scale cuts described in Sec. III F.

1. Linear bias

We parametrize the redshift dependence of the linear bias amplitude b_g (for the lens galaxies, distinct from $b_{g,s}$ for the source sample) by assigning an independent parameter b_g^i to each tomographic redshift bin, avoiding any assumptions of continuous evolution. Similar to the intrinsic alignment parameters, we include these bias parameters by joint LHS according to our analysis priors in Table I.

Given the linear bias, we then build the noiseless galaxy count map of tomographic bin i as

$$\hat{n}_g^i = \langle n_g^i \rangle \left(1 + b_g^i \frac{\delta_m^i - \langle \delta_m^i \rangle}{\langle \delta_m^i \rangle} \right), \tag{9}$$

where $\langle n_g \rangle$ is the mean number of lens galaxies per pixel. For large biases in regions of strong underdensity, the expression in Eq. (9) can produce unphysical negative values. We address this by truncating negative pixels to zero and renormalizing the map to preserve the total galaxy count, an approach detailed in [121].

In Sec. VI A, we validate this simple linear prescription against more sophisticated galaxy clustering models for appropriate scale cuts. Specifically, we condition the inference pipeline on external simulations described in Sec. V B, which employ subhalo abundance matching and halo occupation distribution models.

2. Poisson noise

So far, the maps contain only the pure galaxy clustering signal according to our linear prescription. To add shot noise, we follow K22 and replace each pixel value by a draw from an independent Poisson distribution with a mean equal to the noiseless prediction of the number of galaxies,

$$n_g^{\mathrm{pix}} = \mathrm{Poisson}[\hat{n}_g^{\mathrm{pix}}].$$

Within the pipeline, Poisson noise plays an analogous role to shape noise for weak lensing.

F. Map-level smoothing and scale cuts

Our simulation-based inference approach uses map-level compression networks that operate directly in real space (Sec. IV A 2). Consequently, we must apply scale cuts to the real-space maps. This contrasts with summary statistics like the binned power spectrum (Sec. IV A 3), where scales can be excluded retroactively from the data vector.

1. Small scales

Several assumptions in our forward model (such as the linear bias prescription for galaxy clustering or the effective baryonification) are expected to break down at sufficiently small scales. To mitigate this source of model misspecification, we follow the standard DES strategy of excluding affected scales from the analysis.

We implement this removal of small-scale information through a two-step process detailed in Appendix A. First, we smooth the maps by convolving them with a Gaussian kernel of scale defined by its standard deviation σ. Second, we add a small amount of pixelwise white noise after smoothing to ensure that small-scale information is removed rather than merely suppressed, yielding an approximate maximal multipole $\tilde{\ell}_{\max}$ (defined in Fig. 17).

To determine the extent of the smoothing kernel, we follow [5,68] and fix a comoving transverse scale at the fiducial cosmology. From this, we compute distinct smoothing scales $\sigma_{\min}^i$ per redshift bin i, where the varying mean radial distances of the redshift distributions $n^i(z)$ translate the fixed transverse scale into different angular smoothing scales. Since we keep these angular smoothing scales constant for all cosmologies on the parameter grid, they correspond to slightly varying transverse scales for different cosmologies.

For weak lensing and galaxy clustering maps, respectively, we set our fiducial smoothing scales using comoving transverse scales R of

$$R = 8\ \mathrm{Mpc/h} \Rightarrow \sigma^i_{\mathrm{min}} = [12.5, 8.5, 6.4, 5.4]\ \mathrm{arc\ min},$$
$$R = 32\ \mathrm{Mpc/h} \Rightarrow \sigma^i_{\mathrm{min}} = [55.3, 37.8, 28.9, 24.2]\ \mathrm{arc\ min}, \quad (10)$$

which are determined by the modeling-robustness tests in Sec. VI A using external simulations.

2. Large scales

We implement a high-pass filter to suppress supersurvey modes that could induce unwanted correlations between the four footprints described in Sec. III C. Following [26], this filter applies a hard cut in harmonic space, setting $a_{\ell m} = 0$ for all $\ell < \ell_{\mathrm{min}} = 30$. The $a_{\ell m}$ coefficients arise from decomposing the map m into spherical harmonics $Y_{\ell m}$ up to order ℓ_{max},

$$m(\theta, \phi) \approx \sum_{\ell=0}^{\ell_{\mathrm{max}}} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\theta, \phi). \quad (11)$$

An example map resulting from the application of the low- and high-pass filters followed by the addition of white noise is compared to the original tomographic map in Fig. 6. All maps use resolution $n_{\mathrm{side}} = 512$.

G. Dataset size

In total, we generate $40\,000(= 1000\ \mathrm{permutations} \times 4\ \mathrm{footprint\ cutouts} \times 10\ \mathrm{noise\ realizations})$ semi-independent realizations at the fiducial cosmology and $400(= 20\ \mathrm{permutations} \times 4\ \mathrm{footprint\ cutouts} \times 5\ \mathrm{noise\ realizations})$ per grid cosmology, yielding $1\,000\,000(= 400 \times 2500\ \mathrm{cosmologies})$ grid maps overall. This results in a total dataset size of more than 15 TB.

IV. SIMULATION-BASED INFERENCE

Bayes' theorem [122] relates the posterior $p(\theta|x)$ to the likelihood $p(x|\theta)$ and prior $p(\theta)$ as

$$p(\theta|x) \propto p(x|\theta) p(\theta), \quad (12)$$

where θ denotes the model parameters and x the data, neglecting the constant Bayesian evidence $p(x)$ in the denominator.

Traditional cosmological parameter inference relies on analytical or semianalytical models to connect theoretical predictions with observational data (e.g. [7]). However, the increasing complexity of cosmological and observational models can render these likelihoods intractable or computationally prohibitive. In weak lensing and galaxy clustering cosmology, such intractability commonly arises from the inapplicability of the central limit theorem [123,124], nonlinear structure formation at late times [123,125], realistic measurement systematics [126–128], and summary statistics such as the map-level compression networks employed in this work, for which even the likelihood's

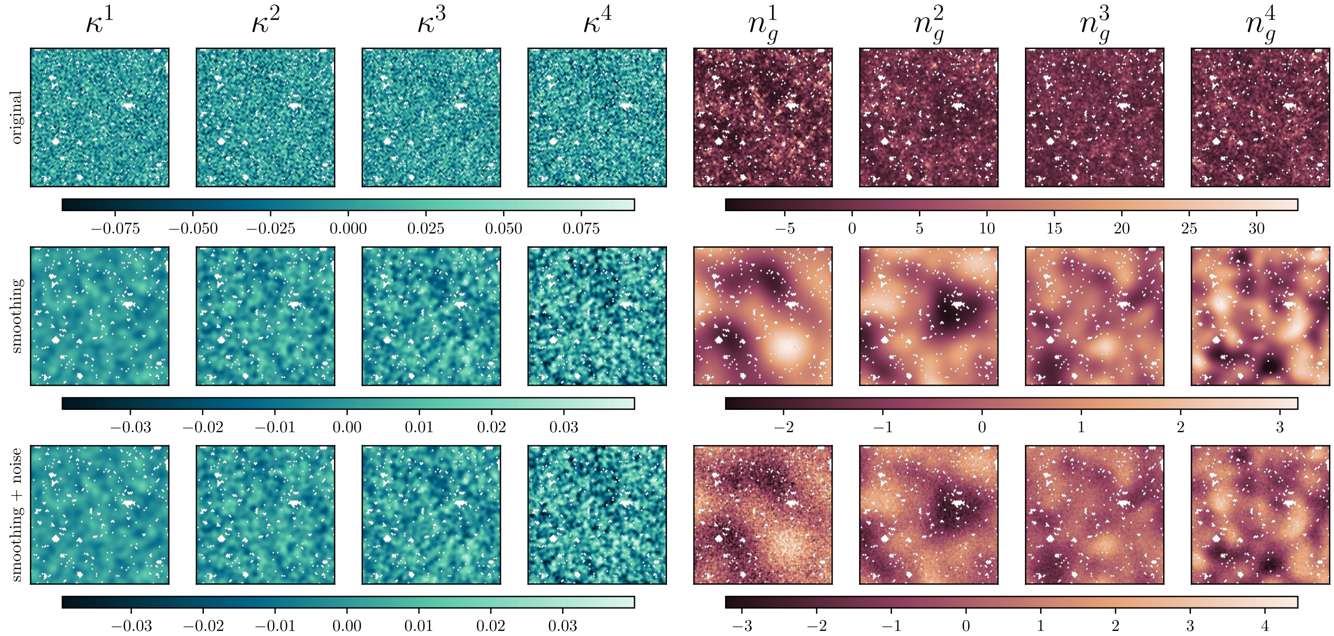

FIG. 6. Comparison of the original unsmoothed maps (upper row) and the smoothed maps obtained following the steps described in Sec. III F (middle and lower rows). The angular smoothing scale is derived from a fixed comoving transverse scale, resulting in varying smoothing between the tomographic redshift bins (columns): Bins closer in redshift to the observer (lower index) are smoothed more than those further away. The squares show gnomic projections of size 10° × 10°, the color bars are derived from the 0.001 and 0.999 quantiles, and the masking from Fig. 4 is visible in white.

expectation value cannot be predicted analytically and must instead be estimated from simulations.

In such scenarios, SBI (also known as implicit or likelihood-free inference) offers a powerful alternative by deriving posterior constraints $p(\theta|x_{\mathrm{obs}})$ conditioned on an observation x_{obs} without the need for explicit likelihood evaluation [49].

A common strategy for SBI is splitting the problem into two consecutive steps: First, a compression function is constructed to map the data x to a lower-dimensional space, facilitating subsequent density estimation. In this work, this mapping is parametrized as an artificial neural network taking the forward-modeled maps or power spectra as input. Second, a conditional probability distribution—such as the likelihood or posterior—relating the compressed data to the parameters θ is learned from an ensemble of simulations using neural density estimation. Here, the density estimator is implemented as a normalizing flow and approximates the unknown likelihood. This completes the framework, allowing amortized inference of the desired posterior constraints conditioned on any (mock) observation.

In Sec. IV A, we introduce the optimization objective that allows us to train neural networks to implement informative compression functions. Section IV A 2 details how we train map-level compression networks, which have access to the whole information content of the forward-modeled maps in principle. Following previous works [F22, [8,27,37,38,41–44,46]], we present angular power spectra as a second-order baseline to compare to the map level in Sec. IV A 3. Consistently following a simulation-based approach, we compute these angular power spectra from our forward-modeled maps.

A. Neural compression

The forward-modeled maps in this work contain millions of pixels, and although the power spectrum detailed in Sec. IV A 3 constitutes a summary statistic of the maps by itself, the resulting data vector still has over a thousand entries. With such high dimensionality, estimating the likelihood function becomes practically intractable due to the curse of dimensionality. This motivates the introduction of a "compression function"

$$\begin{aligned} S: \mathbb{R}^{d_x} &\to \mathbb{R}^{d_s} \\ x &\mapsto S(x) \end{aligned}$$

that maps from the high-dimensional input space of maps or power spectra to a much lower dimensionality $d_s \ll d_x$ where density estimation is feasible. In our case, d_s is of the same order as the number of constrained parameters. The output of the compression function is also referred to as a summary statistic.

Ideally, this compression preserves all information about the parameter vector θ contained in the input x. Such a compression is called sufficient (in the Bayesian sense [129]), meaning the posterior distribution remains unchanged when conditioned on the compression rather than the full input,

$$p(\theta|x) = p(\theta|S(x)). \tag{13}$$

While no theoretical guarantee of sufficiency exists for our setting, this is not strictly necessary. Suboptimal compressions yield inflated or overly conservative posterior contours rather than biased ones, since the same compression function is applied consistently to both forward-modeled and observational data [27]. Moreover, studies on simplified weak lensing models have empirically demonstrated that neural compressions S_φ with trainable parameters φ can achieve sufficiency in practice, provided the network operates at the map level and is optimized with respect to an appropriate loss function like the one discussed below [39,48].

The implementation is available on GitHub [130].

1. Optimization objective

Alongside the architecture, the choice of optimization objective and procedure determines the quality of the compression; specifically, how much information the compressed outputs retain about the parameters of interest in relation to Eq. (13).

The simplest loss functions that produce manifestly informative compressions within our setting are the mean squared error (e.g. [27,46]) and mean absolute error (e.g. [42–44]). These objectives train the network to respectively predict the mean or median of the one-dimensional marginal posterior of the individual components θ_i of the parameter vector, which does not fully characterize the multivariate posterior distribution in general and can therefore lead to suboptimal compression [41,48].

Hence, we employ the information theoretical objective dubbed variational mutual information maximization (VMIM, introduced to the field of cosmology in [41]). Note that, in this case, a single network is trained to find a compression that is jointly informative on the whole parameter vector θ.

Under this objective, the mutual information

$$\begin{aligned} I(s,\theta) &= D_{\mathrm{KL}}(p(s,\theta)\|p(s)p(\theta)) \\ &= \mathbb{E}_{p(s,\theta)}[\log p(\theta|s)] - H(\theta) \end{aligned} \tag{14}$$

between the output of the trainable compression $s \coloneqq S_\varphi(x)$ and the parameters θ is maximized, where D_{KL} is the Kullback-Leibler (KL) divergence [131] and H the Shannon entropy [132]. This expression is intractable because the exact posterior $p(\theta|s)$ is unknown.

However, there exist a number of tractable lower bounds to Eq. (14) that can instead serve as optimization targets (see e.g. [133]). Following [41,48], and based on our empirical finding that this approach outperforms alternative estimators, we employ the variational lower bound [134]

$$I(s,\theta) \geq \mathbb{E}_{p(s,\theta)}[\log q(\theta|s;\phi)] - H(\theta),$$

where we have introduced the variational distribution $q(\theta|s;\phi)$ with trainable parameters ϕ to approximate the unknown true posterior $p(\theta|s)$. The approximate learning objective then becomes

$$\arg\max_{\varphi,\phi} \mathbb{E}_{p(s,\theta)}[\log q(\theta|s;\phi)],$$

where we have discarded the constant entropy term. This yields the loss function

$$L_{\mathrm{VMIM}}(x,\theta;\varphi,\phi) = -\log q(\theta|S_\varphi(x);\phi), \qquad (15)$$

which we estimate from minibatches following standard practice.

In our multiprobe setup, the parameter vector $\theta = (\Omega_m, \sigma_8, w, A_{\mathrm{IA}}, \eta_{A_{\mathrm{IA}}}, b_{\mathrm{TA}}, b_g^1 - b_g^4)$ comprises all ten parameters to be constrained (see the green ellipses in Fig. 2), while marginalized nuisance parameters do not enter the loss. For simplicity, we implement the density estimator q_ϕ as a Gaussian mixture model.

Note that, after training, we discard the variational distribution $q(\theta|s;\phi)$ and retain only the compression network S_{φ^*} with optimal parameters φ^*. We discard the approximate posterior q_ϕ for two reasons: First, the prior implicit in the parameter space sampling by the COSMOGRIDV1 suite consists of a wide and tight Sobol sequence (see Fig. 3). When learning the posterior q_ϕ directly as in Eq. (15) (neural posterior estimation [49,135–137]), this nonuniform prior enters the posterior implicitly. Learning the likelihood instead (neural likelihood estimation [52,53,138]) avoids this issue and allows us to use a simpler uniform prior. Second, dividing the problem into two consecutive steps by fitting a dedicated density estimator for a fixed compression network enables more accurate density estimation in practice. For more details, see Sec. IV B.

2. Map level

The compression networks operating at the map level have, in principle, access to the full information content of the forward-modeled pixelized fields, including non-Gaussian information absent from two-point statistics. We therefore expect these networks to capture additional information beyond what is encoded in the power spectrum, leading to tighter posterior constraints. Depending on the scales considered, this additional information can potentially break parameter degeneracies (e.g. K22).

The input to the networks consists of partial-sky maps within the padded footprint shown in Fig. 4. We treat the two probes and their respective four tomographic redshift bins analogous to color channels (e.g. RGB) in natural images. Hence, for the combined-probe analysis, all eight maps (columns in Fig. 6) are used as input, while probe-specific analyses use only the corresponding four κ or n_g maps. In this setup, all of the tomographic bins are treated symmetrically and can interact freely, allowing the network to capture arbitrary (nonlinear) cross-bin and cross-probe interactions between any number of bins. This is in contrast to the two-point-level compression, which by construction only operates on cross-maps composed of two tomographic bins at a time [see Eq. (16)].

(a) Architecture: All the map-level networks considered in this paper share the same basic structure: feature-extraction layers followed by a regression head, whose design remains unchanged throughout.

Following F22 and [39], our fiducial choice of architecture is the TENSORFLOW [139] implementation of the DEEPSPHERE [140] graph convolutional neural network developed in [112,113]. In this design, the pixels making up the intricate footprint in Fig. 4 are represented as the nodes of a sparse graph that includes edges between a fixed number of neighbors.

The graph construction avoids projecting the survey data from the curved sphere onto a flat image, which invariably introduces distortions that break the symmetry of the spherical geometry and can hinder the learning of convolutional filters. For unmasked inputs, that symmetry is nearly preserved by the DEEPSPHERE graph convolutional layers, which are approximately equivariant under rotations.

As in F22 and [39] and described in Sec. III C, we pad the survey footprint with zeros. This is motivated by our use of the DEEPSPHERE pseudoconvolution layers introduced in [39], which downsample the n_{side} of the internal feature maps according to the hierarchy inherent to the HEALPIX pixelization scheme and require full superpixels.

The feature-extraction part of our fiducial architecture summarized in Table III is made up of a Gaussian smoothing layer operating in real space, the aforementioned pseudoconvolutional layers reducing the n_{side} of the internal representation, Chebyshev convolutions with optional residual connections, and layer normalizations [141]. Throughout, we use the rectified linear unit (ReLU) activation function.

We opt for a dense regression head of a single layer and dropout [142] of rate 0.01. We found this to perform better than an equivariant architecture with a fully convolutional regression head. We suspect that this is the case because the masking we apply breaks the spherical symmetry. The output dimensionality, which is a free choice for the VMIM

TABLE III. Architecture of the fiducial DEEPSPHERE graph convolutional neural network used for map-level compression of combined probes. Each row shows the layer type, tensor output shape for a batch of N_b maps, and number of trainable parameters. The network processes partial-sky HEALPIX maps of $n_{\mathrm{side}} = 512$ with $n_{\mathrm{pix}} = 458752$ pixels and eight feature channels corresponding to redshift bins. The residual layer is repeated five times. The model contains 6.87M trainable parameters in total, with output dimensionality equal to twice the number of (potentially weakly) constrained parameters. Networks for individual probe compression share this architecture except for input and output dimensions.

Layer Type	Output Shape	No. Parameters
Input	$(N_b, 458\,752, 8)$	0
Smoothing	$(N_b, 458\,752, 8)$	0
Pseudo convolution	$(N_b, 114\,688, 32)$	1 056
Pseudo convolution	$(N_b, 28\,672, 64)$	8 256
Pseudo convolution	$(N_b, 7\,168, 128)$	32 896
Chebyshev convolution	$(N_b, 7\,168, 256)$	163 840
Layer-normalization	$(N_b, 7\,168, 256)$	512
Pseudo convolution	$(N_b, 1\,792, 256)$	262 400
Chebyshev convolution	$(N_b, 1\,792, 256)$	327 680
Layer-normalization	$(N_b, 1\,792, 256)$	512
Pseudo convolution	$(N_b, 448, 256)$	262 400
Residual layer	$(N_b, 448, 256)$	656 896
	$\vdots$	
Residual layer	$(N_b, 448, 256)$	656 896
Flatten	$(N_b, 114\,688)$	0
Layer-normalization	$(N_b, 114\,688)$	229 376
Fully connected	$(N_b, 20)$	2 293 780

objective, is fixed to twice the length of the parameter vector θ.

(b) Training: Besides the forward-modeling of the multiprobe maps, training the map-level compression networks is the most computationally demanding step of the inference pipeline. To address this, we utilize the GPU nodes of the Perlmutter cluster [143] at National Energy Research Scientific Computing Center [144], each equipped with four NVIDIA A100 GPUs connected via high-speed Nvlink-3.

The individual networks are trained in a fully data-parallel manner, distributing the global batch across multiple GPUs as local batches. We fix the local batch size N_b to 16 unless stated otherwise. Then, the global batch size is determined by the number of nodes we utilize. For our testing, we only distribute the training over a single node's four GPUs using TENSORFLOW 's built-in mirrored strategy. For the main runs, we scale the training across 4 nodes using HOROVOD [145], yielding a global batch size of 256.

For our fiducial training scheme, we use the ADAM optimizer [146] with default momentum parameters and clip the gradients to a global norm of 1.0 to prevent large parameter updates that destabilize training.

Our warm-up schedule linearly increases the learning rate from 10^{-4} to 10^{-3} over the first 5000 steps. Afterward, we keep the learning rate constant as we have found a cosine decay schedule to lead to overfitting. In total, we train the networks for 500 000 steps, which corresponds to 40 epochs and takes approximately 24 wall hours.

Our validation set is composed of noise realizations that are unseen during training and we did not observe any signs of overfitting in the validation loss. We attribute this to both the large size of our training set and the addition of white noise as part of our scale cut implementation, which serves as a regularizer.

3. Two-point level

In this work, the angular power spectra C_ℓ serve as a baseline for comparison with the map-level results. As a second-order or two-point statistic, the power spectra capture only the Gaussian component of the maps, discarding any additional non-Gaussian content. While the angular power spectrum is a summary statistic of the forward-modeled maps, the dimensionality of the resulting data vector is still too large for direct neural density estimation in our implementation. Therefore, in analogy to the map level, we train a neural network to compress the information on θ contained in the C_ℓ to lower dimensionality. In summary, operating on the two-point-level can be seen as introducing a fixed, intermediate compression step that lacks trainable parameters and is known to be insufficient in the sense of Eq. (13).

We define the pseudo-C_ℓ of tomographic bins i and j using the spherical harmonics coefficients $a_{\ell m}$ from the map decomposition in Eq. (11) as

$$C_\ell^{ij} = \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} \left| \sqrt{a^i_{\ell m}} \sqrt{a^j_{\ell m}} \right|^2, \tag{16}$$

where cross-probe combinations are included if applicable. We compute the $a_{\ell m}$ coefficients from the forward-modeled HEALPIX maps using the decomposition implemented in the HEALPY [147] package. Thus, we follow a simulation-based approach for the C_ℓ too and do not rely on direct theory predictions, which allows us to easily include all of the systematics going into the map-level analysis and ensures direct comparability between the two approaches.

To form data vectors, we average the C_ℓ within 32 square-root spaced bins between $\ell_{\mathrm{min}} = 0$ and $\ell_{\mathrm{max}} = 3n_{\mathrm{side}}$. To be consistent with the map level, the scale cut is applied as a combination of smoothing and white noise as described in Sec. III F and Appendix A, such that the signal is suppressed for some $\tilde{\ell}_{\mathrm{max}} \leq \ell_{\mathrm{max}}$. The fiducial smoothing scales in Eq. (10) yield

$$R = 8\ \mathrm{Mpc/h} \Rightarrow \tilde{\ell}_{\max} = [589, 863, 1159, 1382],$$
$$R = 32\ \mathrm{Mpc/h} \Rightarrow \tilde{\ell}_{\max} = [133, 195, 255, 305]$$

for weak lensing and galaxy clustering, respectively. For comparison, previous simulation-based DES Y3 weak lensing analyses using the Gower Street simulation suite [27] employed a hard cut at $\ell_{\max} = 1024$ for all four METACAL redshift bins [27,30,31,36], while the theory-based DES Y3 harmonic space analysis [148] used scale cuts at $\ell_{\max} = [105, 154, 199, 237]$ for galaxy clustering and $\ell_{\max} = [139, 204, 264, 315]$ for galaxy-galaxy lensing.

We concatenate the binned C_ℓ along the tomographic axis including cross-z and potentially cross-probe bins with $i \neq j$, resulting in a (32×10)-dimensional data vector for the single-probe setting and a (32×36)-dimensional data vector for the probe combination.

(a) Architecture: Based on the results in Appendix D of K22 and our own testing, we implement the compression using a fully connected network, as this simple architecture performs on par with more sophisticated ones for this task.

We define the fiducial network as a layer normalization [141] layer right after the input, followed by two blocks, each containing a dense layer (1024 units and ReLU activation) and dropout (rate of 0.1) in that order. Like for the map-level networks, the final dense output layer has a dimensionality of $2\dim(\theta)$.

(b) Training: Since the data volume is smaller by more than 3 orders of magnitude, training the compression networks at the two-point level is far less computationally demanding than at the map level and can be completed on a single A100 GPU in under half an hour.

For direct comparability with the map-level analysis, we employ the same VMIM training objective here. Similarly, we train the networks with the ADAM optimizer and clip the global gradient norm to 1.0. We perform 300 000 training steps at a batch size of 4096, which is feasible due to the comparatively small size of the binned C_ℓ data vector.

Before feeding the C_ℓ into the networks, we take the logarithm of their absolute value to reduce their dynamic range and improve numerical stability.

B. Neural likelihood estimation

The forward model described in Sec. III generates samples $x \sim p(x|\theta)p(\theta) = p(x, \theta)$, where the vector θ consists of the cosmological parameters entering the N-body simulation, as well as the astrophysical and nuisance parameters incorporated during postprocessing (see Table I and Fig. 2).

In the following, we denote by θ the parameters to be constrained, while the rest of the parameters are marginalized. For the grid subset of the COSMOGRIDV1, this yields pairs $\{x_i, \theta_i\}_{i=1}^N$ with $N = 1\,000\,000$. As discussed in the previous section, the high dimensionality of the data x renders direct density estimation impractical. Therefore, in the following, we consider $\mathcal{D} \coloneqq \{s_i, \theta_i\}_{i=1}^N$, where $s_i = S_{\varphi^*}(x_i)$ is the output of the fixed compression function resulting from training.

The main distinction between different SBI methodologies lies in which probability density they approximate from these samples. In this work, we choose neural likelihood estimation [52,53,138], which learns the inaccessible density $p(s|\theta)$ from the samples $\mathcal{D}$ using a density estimator $q(s|\theta; \phi)$ with trainable parameters ϕ. This way, we avoid making assumptions about the specific functional form of the underlying true likelihood, which is unknown, and we are free to use a different prior than the one implicit in the parameter space sampling of the COSMOGRIDV1 simulations.

Code available on GitHub [149].

1. Normalizing flows

We implement the neural density estimators using normalizing flows (NFs), which model complicated probability distributions by learning a bijection to a simple, for example, Gaussian, base distribution. This bijective mapping is constructed as the composition of multiple discrete layers. For a pedagogical introduction to NFs, we refer the interested reader to [51].

(a) Architecture: We use the FLOWCONDUCTOR [150] package since it offers a wide selection of conditional transformations.

Specifically, we compose the learnable bijection in our NFs from four blocks each made up of a conditional sum-of-sigmoids layer (introduced in Appendix A.1 of [151]) and singular value decomposition layer. The former parametrize a monotonic, elementwise function as the sum-of-sigmoid activations and apply it autoregressively [152], while the latter allow for interactions between dimensions that are absent from those elementwise operations.

(b) Training: In our setting, we only have access to the samples $\mathcal{D}$ and not their underlying probability density. To approximate this unknown true distribution, we minimize the forward KL divergence between the true and approximate distributions, which yields the objective

$$\begin{aligned} &\arg\min_{\phi} D_{KL}(p(s|\theta) \| q(s|\theta; \phi)) \\ &= \arg\max_{\phi} \mathbb{E}_{p(s|\theta)}[\log q(s|\theta; \phi)], \end{aligned} \tag{17}$$

where $q(s|\theta; \phi)$ is the NF with trainable parameters ϕ. By dropping the terms constant with respect to ϕ in the second line, we see that this objective reduces to maximum likelihood estimation [51]. This defines the loss function

$$L_{\mathrm{NF}}(\phi) = -\log q(s|\theta; \phi),$$

which we estimate from minibatches drawn from $\mathcal{D}$, thereby obtaining a Monte Carlo approximation of the expectation in Eq. (17).

Due to the low dimensionality of the compressed data s, training the NFs is comparatively lightweight and fits on a single GPU. For our fiducial setup, we use a batch size of 4096, learning rate of 10^{-4}, gradient clipping to 1.0, cosine decay over 200 epochs, and again the ADAM optimizer.

The quality of the approximation $q(s|\theta;\phi) \approx p(s|\theta)$ is assessed with coverage tests in Sec. VI B.

2. *Inference*

To obtain the posterior constraints $p(\theta|s_{\mathrm{obs}})$ for an observation $s_{\mathrm{obs}} = S_{\varphi^*}(x_{\mathrm{obs}})$, we approximate Eq. (12) as

$$\begin{aligned} p(\theta|s) &\propto p(\theta)p(s|\theta) \\ &\approx p(\theta)q(s|\theta;\phi^*), \end{aligned}$$

and sample using the Markov chain Monte Carlo (MCMC) code EMCEE [153] employing 1024 walkers performing 2000 steps each.

V. MOCK OBSERVATIONS

In this section, we describe the construction and key characteristics of synthetic observations used to conduct end-to-end tests of our inference pipeline's robustness against differing modeling choices.

A. COSMOGRIDV1

The COSMOGRIDV1 simulation suite includes benchmark runs using the alternative N-body simulation settings listed in Table II to assess the adequacy of our fiducial choices for particle number, redshift shell spacing, and replicated box size. These tests address concerns raised by studies such as [154], which demonstrated that varying particle counts can bias SBI when spurious small-scale information is not properly discarded. Since benchmark runs share identical initial conditions with their fiducial counterparts, cosmic variance is eliminated in direct comparisons at the contour level, isolating the impact of the different simulation settings. Furthermore, we use the forward model detailed in Sec. III to construct the benchmark maps, ensuring that observed differences are exclusively attributable to variations in the underlying N-body simulations.

The COSMOGRIDV1 also enables direct comparisons between mock observations with and without the baryonic corrections described in Sec. III A 3 and between our treatment of intrinsic alignment amplitude redshift evolution in Eq. (8) and the standard NLA model.

B. BUZZARD

Our synthetic BUZZARD [55,155] DES Y3 mock observations are generated from a separate forward model not based on the COSMOGRIDV1 that does not follow the data flow in Fig. 2. Therefore, these mocks probe potential model misspecification and, as such, play a pivotal role in validating our pipeline. The primary validation tests we conduct are cosmological parameter recovery and posterior predictive checks, which enable us to determine robust scale cuts for both the weak lensing and galaxy clustering components of our analysis.

To construct the mocks, we use the BUZZARD V2.0 suite [55], which consists of 18 synthetic galaxy catalogs covering the DES Y3 footprint, of which 15 are available to us. The catalogs are generated from the light cone output of dark-matter-only simulations with halo masses above $\sim 5 \times 10^{12} h^{-1} M_\odot$ at $z \leq 0.32$ and $\sim 10^{13} h^{-1} M_\odot$ up to $z \sim 2$ and are populated with galaxies using the ADDGALS method [156]. This results in independent galaxy catalogs with up to $\sim 1.5 \times 10^9$ galaxies.

ADDGALS employs a hybrid approach combining subhalo abundance matching (SHAM) [157] and a halo occupation distribution (HOD) [158] for the galaxy-halo connection. A machine learning model is trained on a calibrated SHAM model to assign central galaxies, while a probability distribution calibrated on SHAM is used to populate the simulation with subhalos. Hence, the procedure employed by ADDGALS differs fundamentally from the linear bias matter-galaxy connection implemented in our COSMOGRIDV1 forward model, which is key to our validation tests.

Additionally, the CALCLENS algorithm [159] is used to compute gravitational shear at the position of each galaxy in the catalog.

For every catalog, we produce a single self-consistent tomographic map of synthetic weak lensing and galaxy clustering observations matching the DES Y3 data in terms of masking, redshift distribution, and average galaxy number counts.

Previously, BUZZARDv2.0 catalogs were used to validate the DES Y3 3×2 pt analysis [55], while older versions of the catalogs [155] have been used to validate cosmological parameter estimation for DES Y1 [160]. The underlying cosmological parameters of the BUZZARD simulations are $\Omega_m = 0.286$, $\sigma_8 = 0.82$, $h = 0.7$, $n_s = 0.96$, and $\Omega_b = 0.046$. The simulations do not model baryonic feedback.

1. *Synthetic source galaxy catalog*

To generate weak lensing mock observations from BUZZARD galaxy catalogs that resemble the DES Y3 source galaxy sample introduced in Sec. II B, we assign observational properties such as signal-to-noise ratio, size, and observed colors to individual galaxies in the catalogs using the BALROG framework [161], which matches galaxies from the DES deep-field catalog to their multiple injections into the DES wide-field image processing pipeline.

This process involves three steps: First, we use the magnitudes of BUZZARD galaxies in the g, r, i, and z bands to identify their closest counterparts in the DES deep-field catalog. Second, we randomly select one wide-field BALROG injection for each matched deep-field galaxy and assign its METACAL properties to the corresponding galaxy in the original BUZZARD catalog. Finally, after applying this procedure to all galaxies in the catalog, we implement the METACAL selection [61]:

```
flags == 0
snr > 10
snr < 1000
size_ratio > 0.5
T < 10
not (T > 2 and snr < 30)
not  (log10(T)  <  (22.25  -  r)/3.5  and
sqrt(e_1**2 + e_2**2) > 0.8)
18 < I < 23.5
15 < (r,z) < 26
-1.5 < (r-i,z-i) < 4,
```

where SNR is the signal-to-noise ratio, `size_ratio` is the ratio between the size `T` and the size of the point spread function (PSF), and `e_1`, `e_2` denote the galaxy ellipticity components. The cuts yield a larger number of galaxies than the true source galaxy catalog, providing flexibility for further subselection. We use this degree of freedom to fine-tune the mock catalog to more accurately reproduce the original METACAL redshift distribution by adjusting galaxy counts within thin redshift intervals dz to match expected values while maintaining a constant total galaxy count per redshift bin.

Based on this processed sample, we can now generate the mock weak lensing shear maps γ_{WL}. For the shape noise component γ_{SN}, we sample the intrinsic ellipticities $\gamma_{j,\mathrm{SN}}$ and weights w_j from the DES Y3 shape noise catalog $\tilde{\gamma}_{\mathrm{METACAL}}$, consistent with the procedure adopted for the COSMOGRIDV1 in Sec. III D 2. We choose not to include intrinsic alignment effects into the mocks, equivalent to an intrinsic alignment amplitude A_{IA} of zero.

The resulting shear values in a given pixel i are weighted by w_j, where j indexes the galaxies falling into pixel i,

$$\gamma_i^{\mathrm{pix}} = \frac{\sum_j w_j(\gamma_{j,\mathrm{WL}} + \gamma_{j,\mathrm{SN}})}{\sum_j w_j}.$$

To ensure consistency with the COSMOGRIDV1 mocks and following the procedure outlined in Sec. III D 1, we apply identical masking, remove B-modes, and convert the shear maps to convergence maps using the Kaiser-Squires transform, yielding the final tomographic $\kappa_{\mathrm{BUZZARD}}$-maps.

Throughout this paper, we assume the total (per-bin) responses to equal unity in both the COSMOGRIDV1 forward model and the construction of the BUZZARD mocks. We will consider realistic response values in the follow-up companion paper dedicated to the analysis of the actual DES Y3 observations.

TABLE IV. Selected properties of the synthetic METACAL sample constructed from a simulated BUZZARD galaxy catalog.

Bin	N_g	$\langle z \rangle$	n_{eff}	σ_e	$\cap n(z)$	$\tilde{\ell}_{\max}$
Full	100 203 633	0.623	5.545	0.260	…	…
1	24 940 369	0.32	1.463	0.244	0.97	589
2	25 280 310	0.51	1.467	0.262	0.97	863
3	24 891 762	0.74	1.471	0.259	0.97	1159
4	25 091 192	0.93	1.449	0.310	0.97	1382

Table IV summarizes the properties of our synthetic METACAL catalogs. The column N_g gives the total number of galaxies in the full sample and each tomographic bin, while $\langle z \rangle$ denotes the mean redshift of the source galaxies. We define the effective number density n_{eff} and shape variance σ_e following [61]. The overlap metric $\cap n(z)$ quantifies the similarity between our reconstructed galaxy redshift distributions and those of the DES Y3 METACAL sample and is computed as the overlap of the normalized $n(z)$ distributions where 0 indicates completely disjoint distributions and 1 indicates identical distributions.

2. Synthetic lens galaxy catalog

In the following, we outline how we postprocess the BUZZARD galaxy catalogs to generate mock observations resembling the DES Y3 MAGLIM lens galaxy sample described in Sec. II C.

Starting from the original BUZZARD catalogs, we first apply the same i-band magnitude cut $i < 4z + 18$ employed in the MAGLIM sample, where i-band magnitudes are available for each galaxy in the BUZZARD catalogs and z is the photometric redshift estimated using the DNF algorithm [69,70].

Since only one of the BUZZARD catalogs includes DNF redshift estimates, we construct the conditional probability distribution $p(z|z_{\mathrm{true}})$ of DNF redshifts z given the true redshifts from this catalog z_{true} and sample from this distribution to impute the remaining catalogs. This allows us to approximately reconstruct the redshift distributions $n^i_{\mathrm{MAGLIM}}(z)$ for each tomographic bin of the MAGLIM catalog i.

As the last step before binning, we fine-tune our catalog to reconstruct the $n_{\mathrm{MAGLIM}}(z)$ distributions more precisely, following the same approach used for the weak lensing mocks.

Table V presents the basic properties of our synthetic BUZZARD galaxy clustering maps for comparison with the original DES Y3 MAGLIM sample. The table shows the total number of galaxies N_g, redshift ranges and mean redshifts $\langle z \rangle$ for each tomographic bin, galaxy number density n_g (per arc min^2), overlap integral $\langle z \rangle$ between our galaxy redshift distributions and those of the DES Y3 MAGLIM catalog, and the maximum multipole $\tilde{\ell}_{\max}$ assumed in the galaxy clustering analysis.

TABLE V. Selected properties of the synthetic MAGLIM sample constructed from a simulated BUZZARD galaxy catalog.

Bin	N_g	Redshift range	$\langle z \rangle$	$\langle n_g \rangle$	$\cap\, n(z)$	$\tilde{\ell}_{\max}$
Full	7 700 661	0.20–0.85	0.54	0.512	⋯	⋯
1	2 256 056	0.20–0.40	0.30	0.150	1.00	133
2	1 609 242	0.40–0.55	0.46	0.107	1.00	195
3	1 639 381	0.55–0.70	0.62	0.109	1.00	255
4	2 195 982	0.70–0.85	0.77	0.146	1.00	305

For more technical details about the BUZZARD mock observations, see Ref. [162].

VI. VALIDATION

In this section, we present tests to determine appropriate scale cuts and validate our inference pipeline at these scales using the mock observations described in Sec. V. The tests are performed for the map-level compression networks (see Sec. IV A 2), which are sensitive to non-Gaussian information, thereby providing a more stringent validation setting than the two-point statistic baseline (see Sec. IV A 3). We confirmed that the two-point reference satisfies the same validation criteria.

We define the parameter S_8 as

$$S_8 \coloneqq \sigma_8 (\Omega_m / 0.3)^{0.5},$$

which breaks the degeneracy between Ω_m and σ_8 that is inherent to weak lensing measurements.

A. Scale cuts

Particular approximations in the forward model detailed in Sec. III are expected to fail at sufficiently small scales, most notably the assumption of linear galaxy biasing in our galaxy clustering maps. Consequently, implementing appropriate scale cuts that discard parts of the data is essential to ensure unbiased parameter inference. This strategy is consistent with previous DES Y3 analyses, where small-scale information was excluded (e.g. [5,20,68,163]) to mitigate potential systematic biases from, for example, inadequate modeling of baryonic effects.

For this work, we adopt conservative scale cuts that pass the validation tests presented below. We do not optimize these cuts to find the least conservative values that would still pass our tests, deferring such optimization to the forthcoming companion paper.

1. Recovery of BUZZARD cosmology

As described in Sec. V B, our DES Y3-like BUZZARD weak lensing and galaxy clustering mocks are constructed from HOD-modeled galaxy catalogs. This contrasts with the COSMOGRIDV1 mocks used to train the compression networks and perform SBI, which are generated directly from probe maps at the pixel level (see Sec. III B). Furthermore, the N-body simulations underlying the BUZZARD catalogs use a different numerical code with higher-fidelity settings. Therefore, these mocks introduce systematic modeling discrepancies (of some degree) relative to the COSMOGRIDV1 mocks, enabling tests of the inference pipeline's robustness to such misspecification or, equivalently, generalization performance under "covariate" shift. We aim to demonstrate that these differences in input do not significantly impact the results.

As a first validation, we test whether the pipeline correctly recovers the known cosmology of the BUZZARD simulations. Following [160], we evaluate the joint posterior over the ensemble of mocks x_i,

$$p(\theta | \{s(x_i)\}_{i=1}^{N}) \propto p(\theta) \prod_{i=1}^{N} p(s(x_i) | \theta), \qquad (18)$$

where $N = 15$. This expansion assumes conditional independence between the x_i given θ and simulates the constraining power of a survey with N-fold larger area, thereby reducing the impact of cosmic variance. The independence assumption is justified because, analogous to our COSMOGRIDV1 mocks, we remove potential large-scale correlations between mock observations constructed from disjoint sky regions of the same underlying N-body simulation by applying an $\ell_{\min}$-cut as detailed in Sec. III F 2. Furthermore, the respective noise realizations are independent by construction.

We present the results for weak lensing and galaxy clustering individually, and their combination, in Fig. 7. For each configuration, the maximum *a posteriori* (MAP) estimates of $p(\theta | \{s(x_i)\}_{i=1}^{N})$ lie within 0.5σ of the true cosmological parameters $\theta = \{\Omega_m, S_8, w\}$ in the two-dimensional marginals. This demonstrates that potential systematic biases from the distributional mismatch between the COSMOGRIDV1 and BUZZARD simulations are subdominant to statistical uncertainties. We therefore conclude that, for the considered scale cuts, the inference pipeline passes the test.

2. Systematic mock contamination

This section presents additional tests designed to validate the inference pipeline's robustness to modeling discrepancies using mock observations intentionally contaminated with systematic errors. Here, the mocks for posterior inference are generated within the COSMOGRIDV1 forward model while varying only one selected property at a time. Shifts in the posterior constraints are therefore directly attributable to the known modeling difference. We consider tests successful when shifts remain below 0.3σ in the Ω_m-S_8 plane, the standard criterion for DES Y3 (e.g. [5]). Note that this differs from the previous cosmology recovery

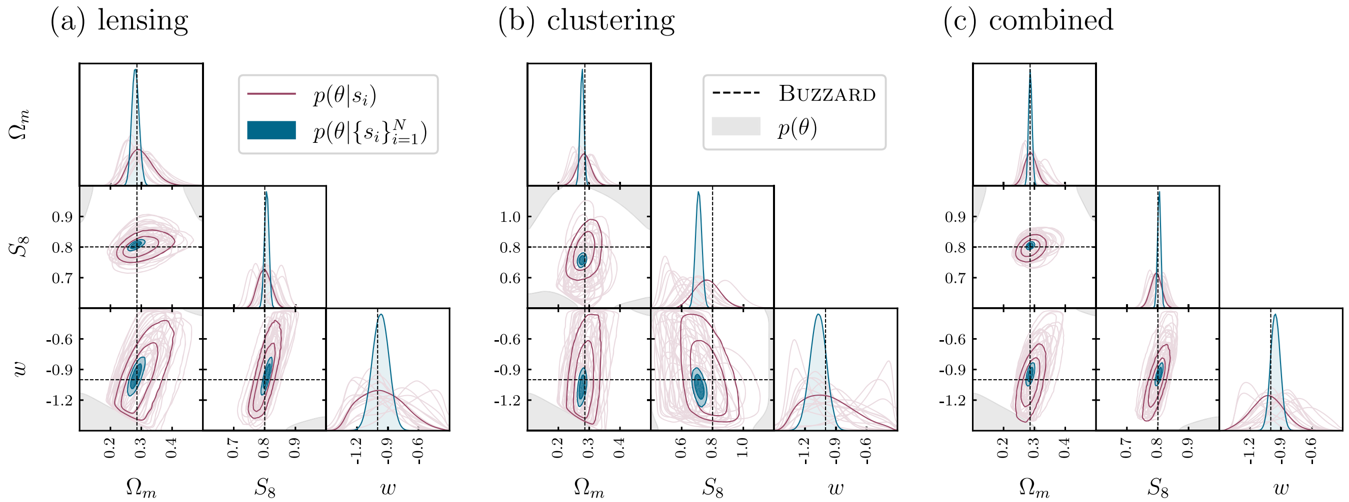

FIG. 7. Cosmological parameter recovery for individual probes [(a) and (b)] and their (c) combination using an ensemble of $N = 15$ BUZZARD mocks x_i with map-level compression $s_i \coloneqq s(x_i)$. In all cases, the MAP estimate of the joint posterior $p(\theta|\{s_i\}_{i=1}^N)$ defined in Eq. (18) (solid blue contours) lies within $< 0.5\sigma$ of a single observation $p(\theta|s_i)$ shifted to align with the ground truth (solid red contour). This result demonstrates that potential systematic biases from modeling differences between the COSMOGRIDV1 and BUZZARD simulations remain below the statistical uncertainty. Faint red contours show individual realizations x_i, contour levels represent 68th and 95th percentiles, and the gray shaded region indicates where the prior $p(\theta)$ is zero.

test: here, shifts are measured relative to the result obtained from the fiducial setup, not the ground truth.

Following [27,30,36], we perform these tests by conditioning on the componentwise mean

$$\bar{s}_j = \frac{1}{N}\sum_{i=1}^{N} s(x_i)_j \tag{19}$$

of the compression vector $s(x_i)$, where j indexes vector components and i runs over $N = 80$ pseudoindependent realizations. This approach is analogous to standard "noise-free" inference.

Figure 8 shows the posterior 0.3σ-level contours from the following systematically contaminated mocks for the map-level combined-probe analysis:

(a) Simulation settings: The COSMOGRIDV1 benchmark mocks (see Secs. III A 2 and V A) validate the fiducial N-body simulation setup by increasing the (i) box size, (ii) particle counts, and (iii) number of redshift shells.

(b) No baryonification: The baryonification model strength is varied throughout the COSMOGRIDV1 according to Sec. III A 3. Since the associated parameters M_c^0 and ν are not expected to be well constrained at our fiducial scale cuts (F22), we marginalize over them. For this test, we omit baryonification postprocessing from the dark matter light cone, unlike the fiducial mock observation used as reference.

(c) Intrinsic alignment amplitude: We compare our per-bin parametrization of A_{IA} redshift evolution in Eq. (8) with the standard NLA formulation, where the amplitude varies per redshift shell.

(d) Source galaxy bias: We test the sensitivity to deviations from the source galaxy bias values introduced in Sec. III D 2. We consider low and high source galaxy clustering biases corresponding to $\pm 2\sigma$ shifts relative to the distributions in Fig. 5: $b_{g,s}^{\mathrm{low}} = [1.0, 0.96, 1.0, 1.0]$ and $b_{g,s}^{\mathrm{high}} = [1.2, 1.2, 1.3, 1.5]$.

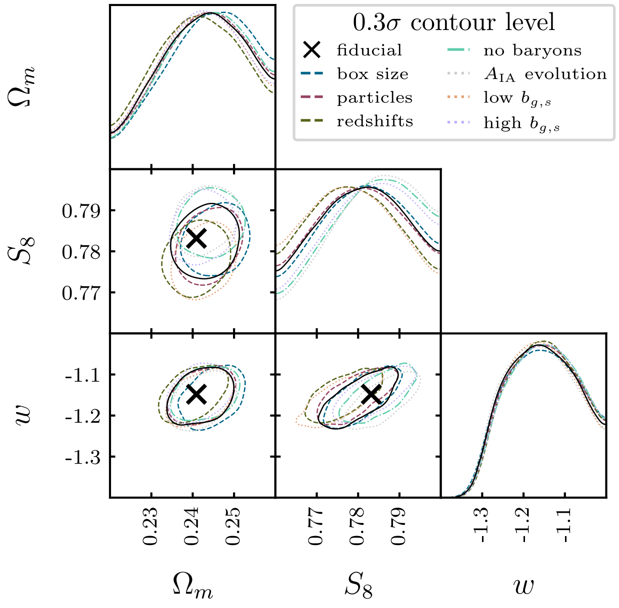

FIG. 8. Impact of systematic error contamination on cosmological constraints inferred from map-level probe combination. The contour lines indicate the 0.3σ tolerance for acceptable shifts within the Ω_m-S_8 plane. All configurations remain within this limit relative to the fiducial MAP (indicated by the cross), demonstrating robustness to alternative modeling.

All configurations produce posterior shifts below 0.3σ relative to the fiducial baseline, passing our robustness criterion for the combined probes analysis. Individual probe analyses (not shown) also satisfy this criterion. The result for the baryonification test indicates that baryonic effects are negligible at our fiducial scale cuts, yet we retain marginalization over the associated parameters as a conservative measure.

3. *Posterior predictive distribution*

The posterior predictive distribution (PPD) [164]

$$p_{\mathrm{PPD}}(x^{\star}|x_{\mathrm{obs}}) = \int p(x^{\star}|\theta)p(\theta|x_{\mathrm{obs}})d\theta \qquad (20)$$

provides an additional diagnostic for detecting potential systematic biases in our inference pipeline by comparing (synthetic) observations x_{obs} with predicted data $x^{\star}$ generated from the fitted model. In this validation test, we employ a BUZZARD mock as x_{obs} while $p(x^{\star}|\theta)$ represents our COSMOGRIDV1 forward model, again creating a scenario where model misspecification may arise. When the forward model adequately captures the underlying data-generating process, samples $x^{\star}$ drawn from the PPD should exhibit statistical properties consistent with the observation x_{obs}. Systematic discrepancies between them indicate deficiencies in the forward model, analogous to biased cosmology recovery or significant posterior shifts due to modeling errors, as examined in previous sections.

We compare the (tomographic cross-) power spectrum of one BUZZARD mock with power spectra from the COSMOGRIDV1 forward model sampled from the PPD using a posterior $p(\theta|x_{\mathrm{obs}})$ derived from map-level galaxy clustering in Fig. 9. The predicted and observed power spectra exhibit good agreement with no systematic deviations exceeding the expected statistical fluctuations. This confirms that our COSMOGRIDV1 -based forward model adequately captures the statistical properties of the BUZZARD simulations at the relevant scales. For brevity, we only show the PPD test for map-level galaxy clustering as it most directly tests the adequacy of our linear galaxy biasing assumption. However, comparable agreement is observed for all other probe configurations.

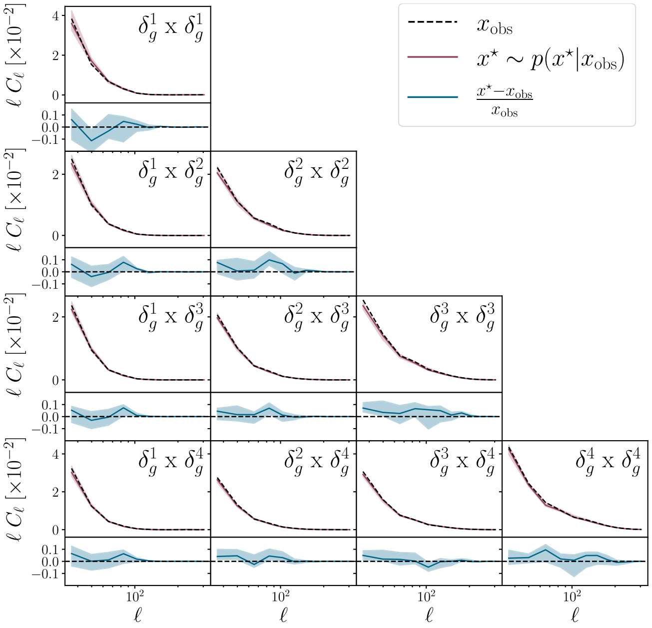

FIG. 9. Comparison between the power spectrum of one BUZZARD simulation (dashed line) to samples from the PPD [see Eq. (20)] derived from map-level galaxy clustering analysis. The solid lines show the mean of the posterior predictive samples, and the shaded regions indicate the 95% credible interval.

B. Posterior coverage tests

In our SBI framework, we parametrize the intractable likelihood as a normalizing flow and sample from the resulting approximate posterior $p(\theta|s(x))$ using an MCMC algorithm, where $s(x)$ is the learned compression of a map x (or power spectrum baseline). For simplicity, we omit s in the following.

We validate this posterior through empirical coverage tests that measure the calibration of the distribution using held-out simulations. Specifically, for a test observation $x_{\mathrm{test}} \sim p(x|\theta_{\mathrm{test}})$ drawn from a known point in parameter space θ_{test}, a well-calibrated posterior $p(\theta|x_{\mathrm{test}})$ should produce nominal $(1-\alpha) \in [0,1]$ credible regions containing θ_{test} with frequency $(1-\alpha)$. Coverage tests therefore involve repeating the inference procedure across many test observations x_{test} to measure this frequency. Although Bayesian credible regions are not theoretically required to satisfy frequentist coverage properties, demonstrating this correspondence empirically indicates that our inference procedure accurately quantifies parameter uncertainty.

We employ two distinct diagnostics to assess calibration: The highest posterior density (HPD) [165] test computes the posterior density at the true parameter value θ_{test} and records the fraction f of posterior samples $\tilde{\theta} \sim p(\theta|x_{\mathrm{test}})$ with lower density. The true parameter is counted as lying inside the nominal $(1-\alpha)$ HPD region whenever $f < 1-\alpha$. Aggregating this indicator over many simulations x_{test} yields a calibration curve, where deviations below (above) the diagonal indicate overconfident (conservative) posteriors.

The tests of accuracy with random points (TARP) [166] method samples random reference points θ_{ref} in parameter space and computes the fraction of posterior samples falling within a ball centered at θ_{ref} and extending to the true value θ_{test}, thereby defining a credible region around the reference point. Repeating this procedure across a large number of reference points and observations x_{test} similarly yields a calibration curve that can be compared to the ideal

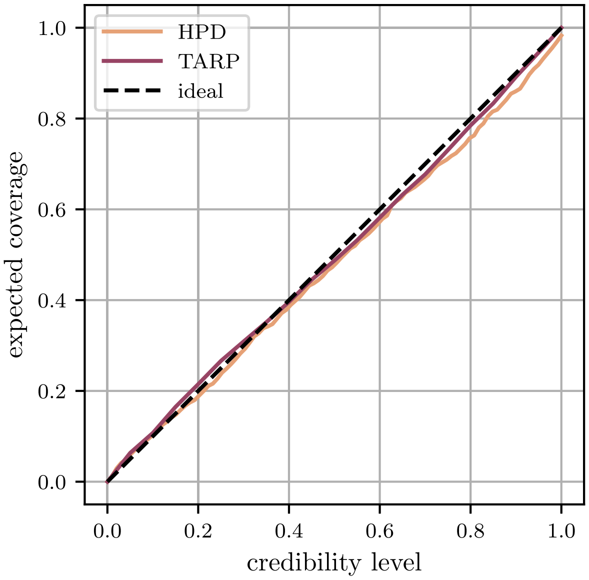

FIG. 10. Empirical validation of posterior coverage through HPD [165] and TARP [166] diagnostics. Curves along the diagonal indicate well-calibrated posteriors where credible regions contain true parameters at their nominal frequencies. Results shown are for the ten-dimensional posterior from map-level combined probes; all other configurations considered in this work perform similarly.

diagonal. We implement TARP using publicly available code [167]. While TARP provides both necessary and sufficient conditions for well-calibrated posteriors, we include HPD as a complementary diagnostic since we have found it to be more sensitive in most cases.

We perform both tests in the full ten-dimensional parameter space encompassing cosmological, intrinsic alignment and galaxy biasing parameters (see Fig. 2), using 1000 synthetic observations x_{test} randomly drawn from held-out COSMOGRIDV1 mocks not seen during training of the compression network or density estimator. The results, presented in Fig. 10, demonstrate accurate calibration.

VII. RESULTS ON MOCK OBSERVATIONS

Having passed extensive validation tests for fiducial scale cuts, we now forecast the constraining power of our posterior inference pipeline by conditioning on the mean compression vector as defined in Eq. (19) over the 15 BUZZARD mocks. Application to the measured DES Y3 METACAL and MAGLIM catalogs, including additional measurement-related systematics modeling and testing, is deferred to a forthcoming companion paper.

We compare performance along two primary axes: first, compression networks operating at the map level versus those using power spectra (our two-point statistic baseline) as input; second, individual probes of weak lensing and galaxy clustering versus their combination. We include additional figures in Appendix B.

A. Map versus two-point level

As explained in Sec. IV A, two-point statistics like the power spectrum only capture Gaussian information. However, the forward-modeled weak lensing and galaxy clustering maps contain additional non-Gaussian information, which is only accessible to the map-level compression networks within our analysis pipeline. Therefore, comparing posterior constraints from map-level and power spectrum analyses provides a lower bound on the non-Gaussian information content of the maps, as extracted by the map-level compression networks.

We quantify the constraining power in the two-dimensional X-Y parameter plane using the figure of merit (FOM) of the marginal distribution [1],

$$\mathrm{FOM}_{X,Y} \coloneqq (\det(\mathrm{Cov}_{X,Y}))^{-0.5}. \tag{21}$$

The results are presented in Table VI and discussed below.

TABLE VI. Marginalized standard deviations $\sigma(\cdot)$ and two-parameter figures of merit (FoM, see Eq. (21)) for different analysis configurations. For map-level statistics, the fractional improvement relative to the power spectrum baseline is shown in parentheses, with positive percentages indicating tighter constraints.

Probe	Statistic	$\sigma(\Omega_m)[\times 100]$	$\sigma(S_8)[\times 100]$	$\sigma(w)[\times 100]$	$\sigma(A_{\mathrm{IA}})[\times 100]$	$\sigma(b_g^1)[\times 100]$	$\mathrm{FoM}_{\Omega_m,S_8}$	$\mathrm{FoM}_{\Omega_m,w}$
Lensing	C_ℓ baseline	5.1	2.9	21	20	⋯	686	92
	map-level	3.9	2.2	21	21	⋯	1340	158
		(+32%)	(+30%)	(+0%)	(−4%)		(+95%)	(+71%)
Clustering	C_ℓ baseline	3.2	16	28	⋯	39	198	121
	map-level	2.3	7.9	24	⋯	16	572	186
		(+35%)	(+105%)	(+17%)		(+147%)	(+189%)	(+54%)
Combined	C_ℓ baseline	3.1	2.7	24	14	13	1224	142
	map-level	2.0	1.8	16	13	6.9	2969	389
		(+54%)	(+52%)	(+52%)	(+4%)	(+89%)	(+143%)	(+175%)

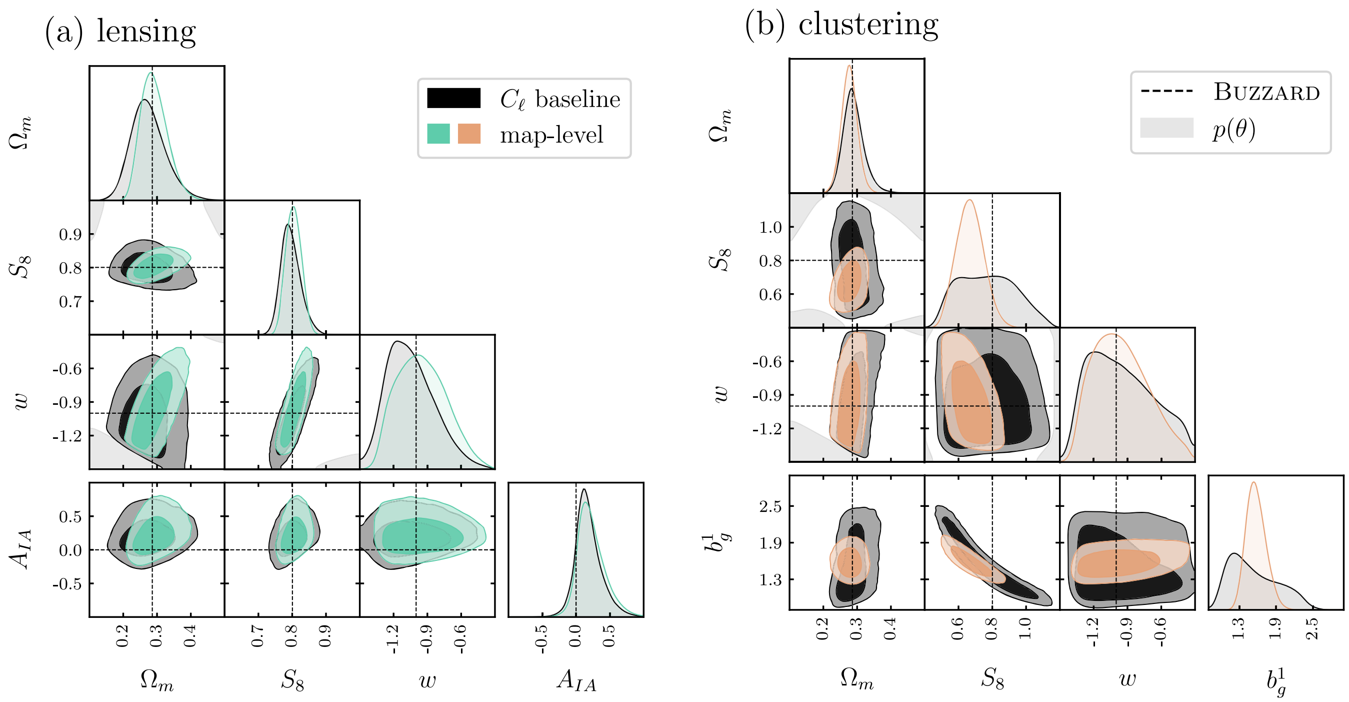

FIG. 11. Comparison of posterior contours between map-level analysis and power spectrum baseline for fiducial scale cuts. The posterior is conditioned on the mean compression of the BUZZARD ensemble. For (a) weak lensing, we exclude the weakly constrained intrinsic alignment parameters $\eta_{A_{\mathrm{IA}}}$ and b_{TA}. For (b) galaxy clustering, we show only b_g^1 as the remaining bias parameters b_g^2–b_g^4 behave similarly. Complete parameter plots are provided in Appendix B.

1. *Weak lensing*

For weak gravitational lensing as an individual probe, we compare two-dimensional posterior marginals in Fig. 11(a). We include the three constrained cosmological parameters $\{\Omega_m, S_8, w\}$ and the intrinsic alignment amplitude A_{IA} in the plot, while excluding the weakly constrained $\eta_{A_{\mathrm{IA}}}$ and b_{TA} parameters for clarity; the complete parameter space is shown in Fig. 19. Map-level networks yield significantly improved constraints compared to the power spectrum baseline, with a 95% increase (i.e. 1.95× improvement) in $\mathrm{FOM}_{\Omega_m,S_8}$. This increase is consistent with previous simulation-based DES Y3 weak lensing analyses employing higher-order summary statistics [26,27,31,36].

In contrast, intrinsic alignment constraints show essentially no improvement. This result differs from K22, which found drastic enhancement in intrinsic alignment constraints for deep-learning-based summaries of weak lensing mass maps compared to conventional two-point statistics. However, the forward models in K22 and this work differ substantially. We employ a three-parameter intrinsic alignment model compared to the vanilla NLA model assumed by K22, constrain wCDM instead of ΛCDM cosmology, and adopt a more realistic stage-III surveylike setup including additional nuisances like redshift errors, multiplicative shear bias, and baryonification. Furthermore, we apply more conservative scale cuts, which reduces the expected gain from higher-order relative to Gaussian statistics.

We demonstrate in Appendix C 1 that the marginal posterior contours of A_{IA} are not centered on the true value of zero due to projection effects. In Appendix C 2, we present posterior results on COSMOGRIDV1 internal mock observations with $A_{\mathrm{IA}} \neq 0$ to show accurate parameter recovery in the presence of intrinsic alignments.

2. *Galaxy clustering*

Analogously, for stand-alone galaxy clustering, we compare posterior results for power spectrum and map-based analyses in Fig. 11(b). For conciseness, we include only a single representative bias parameter b_g^1 and refer to Fig. 20 for the complete parameter space.

The map-level approach partially breaks the degeneracy between linear bias parameters and clustering amplitude σ_8 (or S_8), leading to substantially improved constraints on these parameters despite the conservative scale cuts employed. The improved constraint on S_8 translates to a 189% increase in the FOM in the Ω_m-S_8 plane. For one-dimensional marginal constraints, we find approximately twofold reductions in uncertainty for both S_8 and the b_g^1–b_g^4 parameters, with modest gains for the dark energy parameter w.

To motivate the observed improvement in S_8 and galaxy bias constraints for the map-level analysis, we draw a qualitative analogy with spectroscopic galaxy surveys. Effective field theory (EFT) analyses of the Baryon Oscillation Spectroscopic Survey [168] and DESI [169] have demonstrated that combining the bispectrum with the

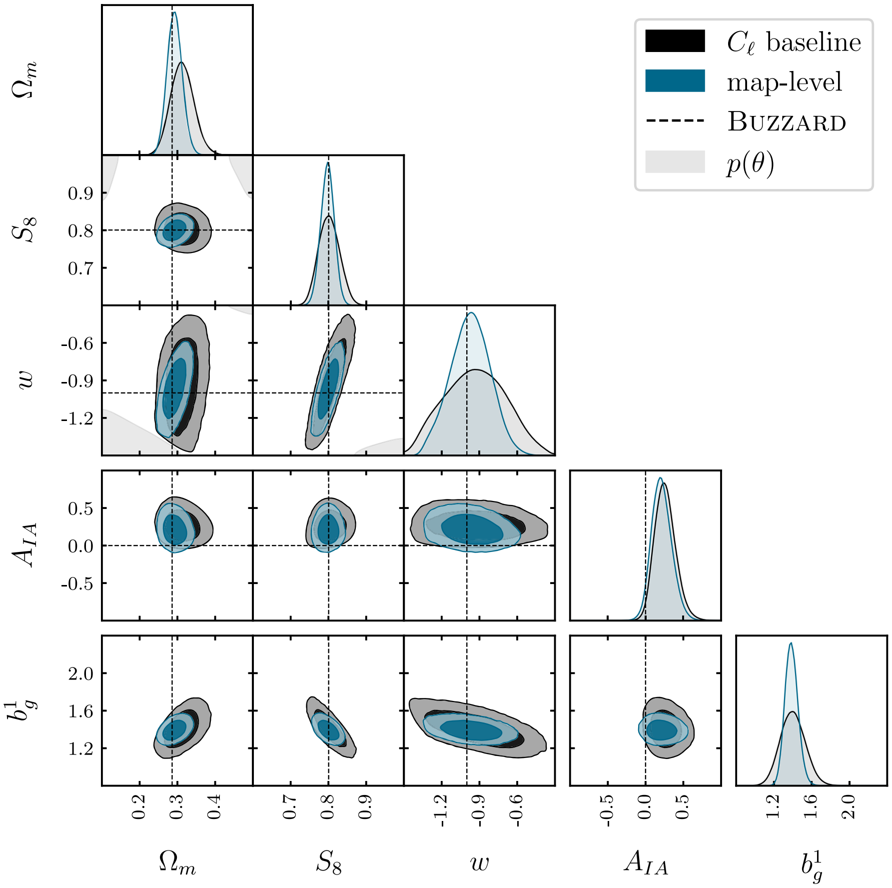

FIG. 12. Like Fig. 11, but for combined probes of weak lensing and galaxy clustering.

power spectrum significantly improves measurement precision for EFT galaxy bias parameters relative to power-spectrum-only analyses, particularly for quadratic terms [170–172].

Considering the EFT linear galaxy bias b_1 (related to but distinct from our b_g parameter), this improvement arises from different degeneracy directions between the power spectrum and bispectrum: As derived in [171], the power spectrum scales quadratically with b_1, while the bispectrum exhibits cubic scaling. These distinct scaling relations create different degeneracy directions relative to σ_8, enabling improved parameter constraints when both statistics are combined.

Since the bispectrum contribution remains statistically significant at our fiducial smoothing scale of 32 Mpc/h for galaxy clustering (Fig. 1 in [171]), and our map-level compression captures the beyond-Gaussian information content of the bispectrum, in principle, we hypothesize that a similar degeneracy-breaking mechanism underlies the improvements observed in our analysis. A quantitative assessment of this plausible analogy is beyond the scope of this work.

3. Combined probes

The main result of this work is the first forecast within DES using combined galaxy clustering and weak lensing maps. For the combined-probe analysis, we compare contours from two-point and map-level statistics in Fig. 12 using the same parameters as in Fig. 11. The complete parameter space is shown in Fig. 21.

Map-based summaries yield significant improvement in constraints on cosmological parameters, reducing marginal

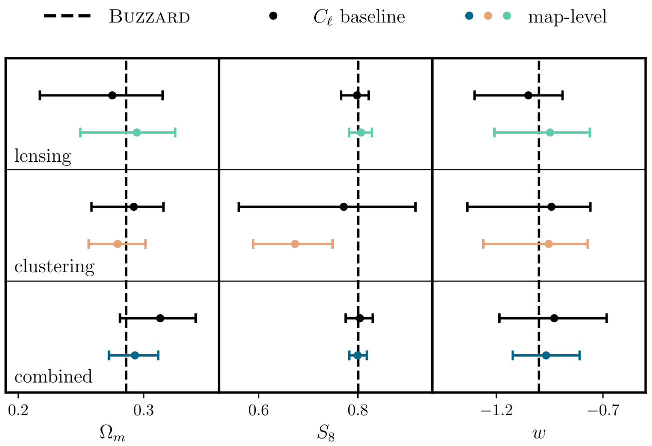

FIG. 13. Comparison of one-dimensional 68% credible intervals for cosmological parameters across different analysis configurations. Dots indicate posterior means.

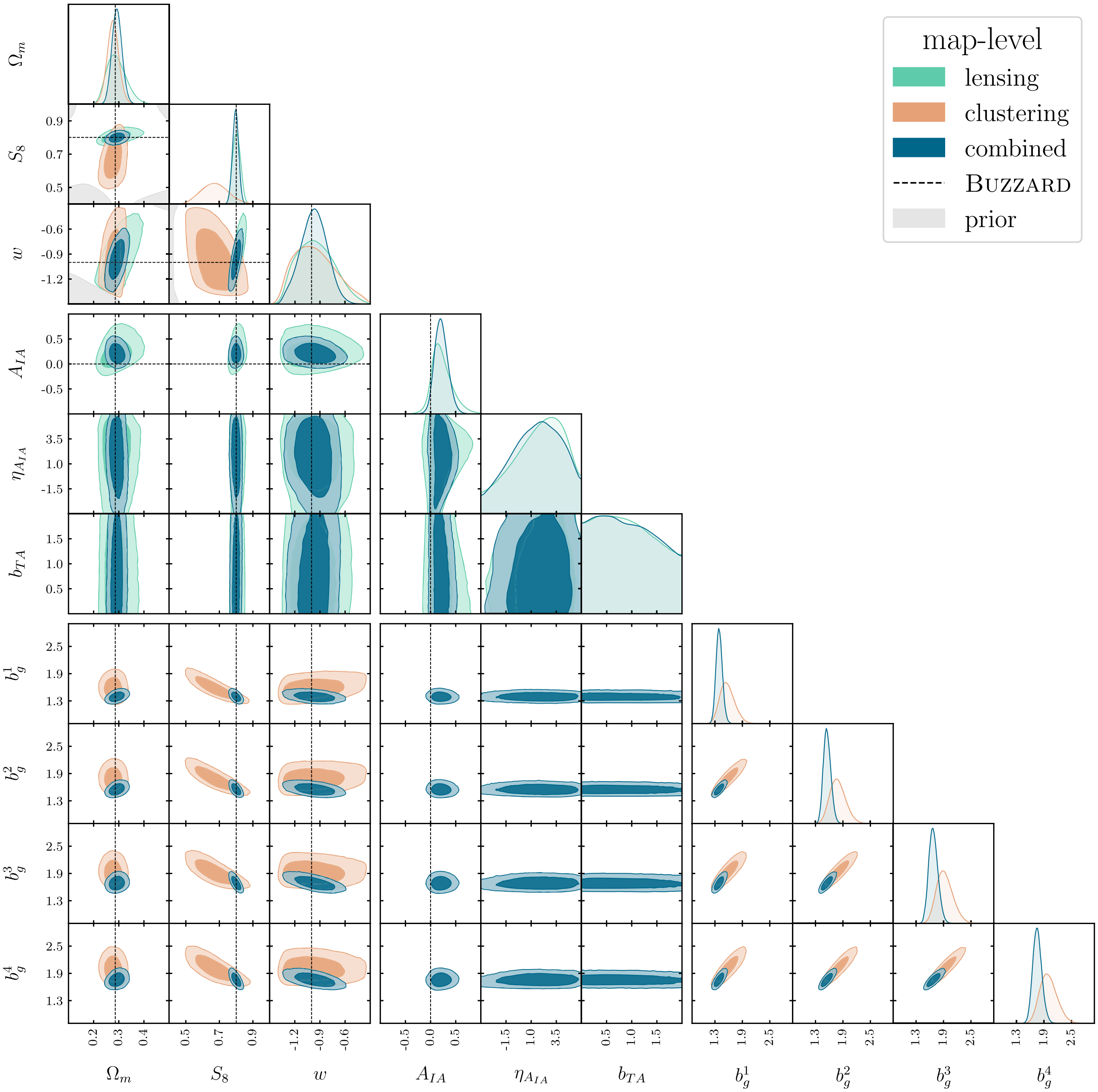

FIG. 14. Comparison of posterior distributions resulting from individual or combined probes for map-level analysis using fiducial scale cuts. The plot shows the full parameter space including cosmological, intrinsic alignment and galaxy biasing parameters. The synthetic observation is the ensemble mean over the BUZZARD mocks, with ground truth values indicated by dashed lines.

standard deviations $\sigma(\cdot)$ by approximately 50% for Ω_m, S_8, and w. Interestingly, while map-level inference enhances constraints on linear bias parameters (with nearly twofold reduction in one-dimensional marginal uncertainties), the intrinsic alignment constraint remains comparable between two-point and map-level analyses, consistent with our weak-lensing-only results. For the overall $\mathrm{FOM}_{\Omega_m,S_8}$, we forecast a 143% improvement due to the extracted non-Gaussian information.

We summarize these findings as one-dimensional marginals in Fig. 13.

B. Probe comparison

In this section, we discuss our map-level results from the perspective of probe combination. Each probe configuration employs an independent compression network trained on tomographic weak lensing maps κ, galaxy clustering maps n_g, or their joint combination. Figure 14 compares posterior constraints from individual and combined probes across the full parameter space, while Fig. 22 presents equivalent results for the two-point statistic baseline.

As expected, the higher sensitivity of weak lensing to the clustering parameter S_8 helps break the degeneracy between the clustering bias parameters b_g^1–b_g^4 and S_8, tightening constraints on these parameters by more than a factor of 2. Relatedly, galaxy clustering provides additional information via larger sensitivity to Ω_m, yielding approximately twofold reduction in the $\mathrm{FOM}_{\Omega_m,S_8}$ compared to weak lensing alone.

We also find substantial improvements in the Ω_m-w plane due to the different degeneracy directions of these parameters for weak lensing and galaxy clustering. This difference is significantly more pronounced for map-level inference than for the power spectrum baseline: While $\mathrm{FOM}_{\Omega_m,w}$ increases by approximately 17% when combining galaxy clustering with weak lensing compared to galaxy clustering alone for power spectra, the corresponding gain for map-based analysis reaches 109%.

VIII. CONCLUSION

In this paper, we present the first pipeline for simulation-based cosmological inference using neural compressions of combined weak lensing and galaxy clustering maps for a realistic stage-III survey setup. We forecast its constraining power and provide detailed validation on simulations in preparation for analysis of the observed DES Y3 data in a forthcoming companion paper.

The main conclusions of this work can be summarized as follows:

(i) We develop a scalable forward model based on the COSMOGRIDV1 simulation suite, generating one million combined weak lensing and galaxy clustering DES Y3 mock observations, requiring approximately 15 TB of storage. Of these maps, 800000 train the neural compression networks and the remaining 200000 enable neural density estimation of the unknown likelihood. This framework allows for robust SBI in a ten-dimensional parameter space.

(ii) In addition to cosmological wCDM parameters, the forward model varies baryonification, intrinsic alignment, and linear galaxy biasing parameters. Furthermore, we marginalize over photometric redshift uncertainties and multiplicative source galaxy shear biases.

(iii) The graph convolutional neural networks we employ to compress the maps operate on the full survey footprint in spherical geometry, avoiding patchification and projection effects. We train them jointly on all constrained parameters using a theoretically motivated mutual information loss.

(iv) As part of the validation effort, we construct 15 independent synthetic observations from BUZZARD galaxy catalogs. We use these potentially out-of-distribution mocks to determine appropriate scale cuts for both galaxy clustering and weak lensing, ensuring robust cosmology recovery and successful posterior predictive checks.

(v) We carry out additional systematics checks, including tests for alternative source clustering bias values and underlying N-body simulation settings. For all of these, we confirm that resulting posterior shifts in the Ω_m-S_8 plane remain below 0.3σ.

(vi) Our map-level compression networks are highly effective at extracting non-Gaussian information from the weak lensing and galaxy clustering fields for the validated scale cuts. We report substantial improvements in cosmological parameter constraints compared to our baseline two-point statistics, with increases in the figure of merit of up to 189% (i.e. 2.89×) depending on the specific probe and parameters, corresponding to narrower posteriors. For galaxy clustering alone, the map-level analysis partially breaks the σ_8 linear bias degeneracy, which is further reduced when combining with weak lensing as expected. We find no significant gain in constraints on intrinsic alignment parameters.

In this work, we did not systematically optimize scale cuts to identify the least conservative configuration satisfying our validation criteria. Instead, we validated a deliberately conservative choice. Consequently, our results represent a lower bound on the constraining power of the methodology, with scope for further improvement through scale optimization in future analyses.

The potential of the inference framework we introduce extends beyond the implemented forward model. For example, while we include baryonic effects and marginalize over associated parameters, their impact is negligible at our current scale cuts. Since the baryonification model remains valid at smaller scales [173], higher resolution weak lensing analyses should be feasible with the existing simulations. For galaxy clustering, our present modeling is primarily limited by the assumption of simple linear biasing for the matter-galaxy connection. Although our neural compression extracts non-Gaussian information from galaxy clustering maps even at conservative smoothing scales, more sophisticated clustering models could further enhance the results. Possible improvements include incorporating higher-order bias terms or, given access to higher-fidelity halo catalogs, a SHAM-like galaxy-halo connection [174,175] in combination with a galaxy population model [176]. However, the current COSMOGRIDV1 suite does not support these approaches, which will be addressed in future versions.

The stage-IV photometric galaxy surveys LSST [177] and Euclid [178] have recently entered operations and will provide data of substantially higher volume and quality than stage-III surveys like DES. While significant challenges in modeling and simulations for stage-IV data analysis remain, the forward model and inference pipeline we develop provide a solid foundation for SBI analyses of next-generation large-scale structure observations.

The following software packages were used in this work: NUMPY [179], SCIPY [180], TENSORFLOW [181], HOROVOD [182], PYTORCH [183], FLOWCONDUCTOR [151,184], WANDB, HEALPY [84,147], EMCEE [153], and H5PY [185]. Job arrays were submitted with ESUB-EPIPE [26,98,99], and plots were created using MATPLOTLIB [186] and TRIANGLECHAIN [8,187].

ACKNOWLEDGMENTS

A. T. thanks Silvan Fischbacher, Alexander Reeves, Veronika Oehl, Tilman Tröster, and Dominik Zürcher for helpful discussions on cosmology and Marcello Negri and Aurelien Lucchi for valuable input on deep learning. We also thank the reviewers for their insightful comments and helpful feedback. This work was supported in part at ETH Zurich by the research Grant No. 200021_192243 from the Swiss National Science Foundation. This work was supported by the Alexander von Humboldt Foundation (T. K). This research used resources of the National Energy Research Scientific Computing Center (NERSC), a Department of Energy User Facility using NERSC Awards No. HEP-ERCAP 0031464 and No. DDR-ERCAP 0034644 (AI4Sci@NERSC). Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft, and the Collaborating Institutions in the Dark Energy Survey. The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, NSF NOIRLab, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, Texas A&M University, and the OzDES Membership Consortium. Based in part on observations at NSF Cerro Tololo Inter-American Observatory at NSF NOIRLab (NOIRLab Prop. ID 2012B-0001; PI: J. Frieman), which is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation. The DES data management system is supported by the National Science Foundation under Grants No. AST-1138766 and No. AST-1536171. The DES participants from Spanish institutions are partially supported by MICINN under Grants No. PID2021-123012, No. PID2021-128989, No. PID2022-141079, No. SEV-2016-0588, No. CEX2020-001058-M, and No. CEX2020-001007-S, some of which include ERDF funds from the European Union. I. F. A. E. is partially funded by the CERCA program of the Generalitat de Catalunya. We acknowledge support from the Brazilian Instituto Nacional de Ciência e Tecnologia (INCT) do e-Universo (CNPq Grant No. 465376/2014-2). This document was prepared by the DES Collaboration using the resources of the Fermi National Accelerator Laboratory (Fermilab), a U.S. Department of Energy, Office of Science, Office of High Energy Physics HEP User Facility. Fermilab is managed by Fermi Forward Discovery Group, LLC, acting under Contract No. 89243024CSC000002. A. T. led the analysis, implemented the forward modeling, neural compression, and simulation-based inference pipelines, and served as primary author of the manuscript. J. B. generated the BUZZARD mocks, co-led the validation, and advised the project. T. K. co-conceptualized the project, coled and provided the COSMOGRIDV1 simulation suite, performed the map projection and baryonification, and advised the project. V. A. helped conceptualize the forward model, interpret results, and provided manuscript feedback. J. F. co-led the COSMOGRIDV1 simulation suite and advised the pipeline development, contributing a precursor implementation. A. R. co-conceptualized the project and advised on all aspects.

A. T.: Led the analysis; implemented the forward modeling, neural compression, and simulation-based inference pipelines; and served as primary author of the manuscript. J. Bucko.: Generated the BUZZARD mocks, co-led the validation, and advised the project. T. K.: Co-conceptualized the project, co-led and provided the BUZZARD simulation suite, performed the map projection and baryonification, and advised the project. V. A.: Helped conceptualize the forward model, interpret results, and provided manuscript feedback. J. F.: Co-led the BUZZARD simulation suite and advised the pipeline development, contributing a precursor implementation. A. Refregier: Co-conceptualized the project and advised on all aspects. A. F., D. A., F. C.: DES internal reviewers. M. G., N. J.: DES beyond-2pt analysis team members. J. DeRose., R. W., M. B., E. R., S. P., N. M., A. Amon, J. Myles.: Production of BUZZARD simulations. J. Myles., A. Alarcon, A. Amon, C. S., S. E., J. DeRose., J. McCullough., D. G., G. B., M. T., S. Dodelson., A. Campos., N. M., B. Yin., M. Raveri: Production of the DES redshift distribution. M. G., E. Sheldon., A. Amon., M. B., M. T., A. Choi., C. Doux., N. M., A. N. A., I. H., D. G., G. B., M. J., L. F. S., A. F., J. McCullough., R. P. R., R. Chen., C. C., S. P., I. T., J. P., J. E. P., C. S.: Production of the DES shape catalog. M. J., G. B., A. Amon., C. Davis., P. F. L., K. B., I. H., M. G., A. Roodman.: Production of the DES PSF. S. E., B. Yanny., N. K., E. H., Y. Z.: Production of the DES Balrog. A. P., M. C., P. F., J. E.: Production of the DES lens galaxy catalog. M. Rodriguez-Monroy., N. W., J. E., M. C.: Production of lens galaxy systematics maps.

DATA AVAILABILITY

The data that support the findings of this article are openly available [60,74,75,130,149].

APPENDIX A: MAP-LEVEL SMOOTHING AND SCALE CUTS

This appendix describes the implementation of the small-scale cuts discussed in Sec. III F 1 and presents validation of the method.

1. Implementation details

We implement the removal of small-scale information in two subsequent steps.

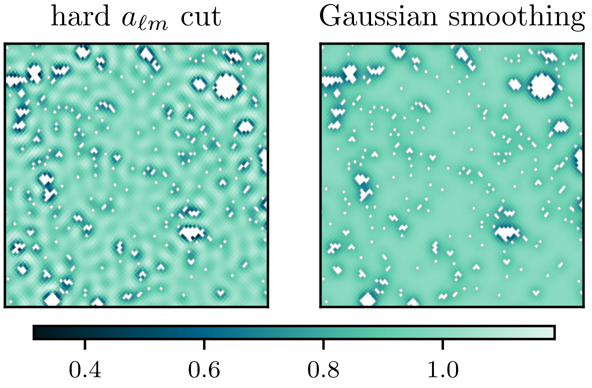

FIG. 15. Comparison between the artifacts introduced by a hard cut in harmonic space and Gaussian smoothing in real space.

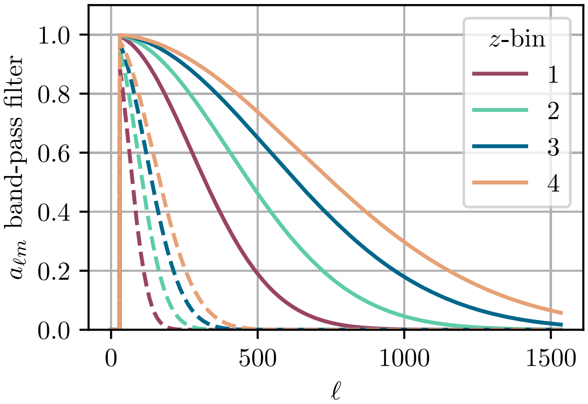

FIG. 16. Bandpass filters in $a_{\ell m}$-space defined by the factor in Eq. (A1) and a hard $\ell_{\min} = 30$ cut for the fiducial smoothing scales used in this work. Solid and dashed lines correspond to the redshift bins of the source and lens galaxy sample, respectively.

(a) Gaussian smoothing: Because the survey footprint depicted in Fig. 4 covers only part of the sky and has a complicated shape, hard cuts in frequency (or spherical harmonics) space introduce ringing artifacts in real space (see Fig. 15), which can hinder the learning of convolutional filters.

Therefore, to remove small scales, we smooth the maps by convolving them with a Gaussian kernel, which mitigates this issue. The smoothing scale is defined by the kernel's standard deviation σ or full width at half maximum ($\mathrm{FWHM} = 2\sqrt{2\ln 2}\sigma$).

A useful property of Gaussian smoothing is its closed analytical form in the frequency domain, where smoothing is accomplished by multiplication of the $a_{\ell m}$-coefficients from the harmonic decomposition in Eq. (11) by the factor

$$c_{\text{low-pass}}(\ell;\sigma) = \exp\left(-\frac{1}{2}\ell(\ell+1)\sigma^2\right), \qquad \text{(A1)}$$

which depends on the multipole order $\ell \in \mathbb{Z}_{>0}$. As ℓ increases, the factor becomes smaller, leading to greater suppression of higher-order modes. The $c_{\text{low-pass}}$ filters resulting from our fiducial scale cuts in Eq. (10) are plotted in Fig. 16.

Beyond reducing artifacts, another advantage of our implementation of Gaussian smoothing over a hard cut in $a_{\ell m}$-space is that we can perform the smoothing operation on the fly, requiring storage of only a single full-resolution version of the dataset.

(b) White noise: Because the Gaussian smoothing kernel (in principle) has infinite support in real and spherical harmonics space, the smoothing operation is theoretically invertible in frequency space by simply dividing the $a_{\ell m}$-coefficients through the factor in Eq. (A1). Thus, mathematically there is no loss of information; the small-scale information is merely suppressed, not removed. While this

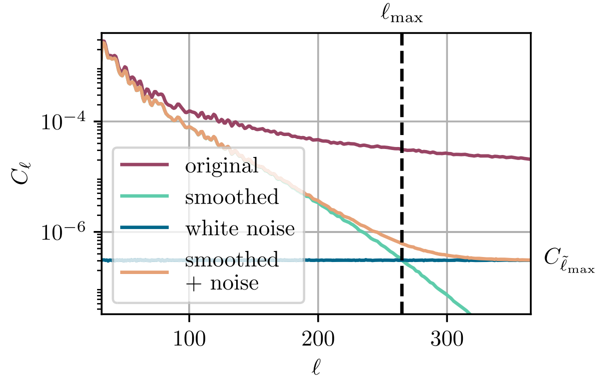

FIG. 17. Illustration of the C_ℓ-level definition for the amount of white noise added after smoothing. Per construction, the "smoothed" C_ℓ (obtained from convolution with a Gaussian kernel) are suppressed to 1% of the original values at $\tilde{\ell}_{\max}$. The "white noise" (generated from independent Gaussian samples) is calibrated to intersect the smoothed C_ℓ at that value denoted $C_{\tilde{\ell}_{\max}}$. This example depicts the first galaxy clustering redshift bin, which undergoes the most smoothing.

is not the case in practice due to the finite precision of floating point numbers, we choose not to rely on such numerical noise. Instead, we address the issue by irreversibly adding white Gaussian noise of a certain scale to the maps after smoothing.

We determine this noise level separately for each tomographic bin i with smoothing scale $\sigma^i_{\min}$ by first finding the smallest spherical harmonic order $\tilde{\ell}^i_{\max}$ for which Eq. (A1) is below a threshold of 1%. In other words (and omitting the bin index for improved readability), $\tilde{\ell}_{\max}$ is the scale at which the smoothing reduces the $a_{\ell m}$-coefficients to 1% of their original value. Then, we take the mean of the angular power spectrum C_ℓ (as defined in Sec. IVA 3) over all realizations at the fiducial cosmology and evaluate the resulting curve at the given $\tilde{\ell}_{\max}$. The obtained value $C_{\tilde{\ell}_{\max}}$ defines the level of white noise we apply. The different power spectra illustrating the steps are plotted in Fig. 17.

We construct white-noise HEALPIX maps by drawing pixelwise samples from independent scalar Gaussians of mean zero and shared standard deviation σ_{noise} (not to be confused with the smoothing scale $\sigma_{\min}$). The standard deviation σ_{noise} is chosen such that the approximately constant power spectrum of the white-noise map matches the noise level $C_{\tilde{\ell}_{\max}}$. The connection between the white-noise power spectrum and the Gaussian distributions is given by

$$C_{\tilde{\ell}_{\max}} = \frac{4\pi f_{\text{sky}}}{n_{\text{pix}}} \sigma^2_{\text{noise}},$$

where f_{sky} is the fraction of the sky occupied by the survey footprint and n_{pix} is the number of pixels contained in the full sky at n_{side} [39].

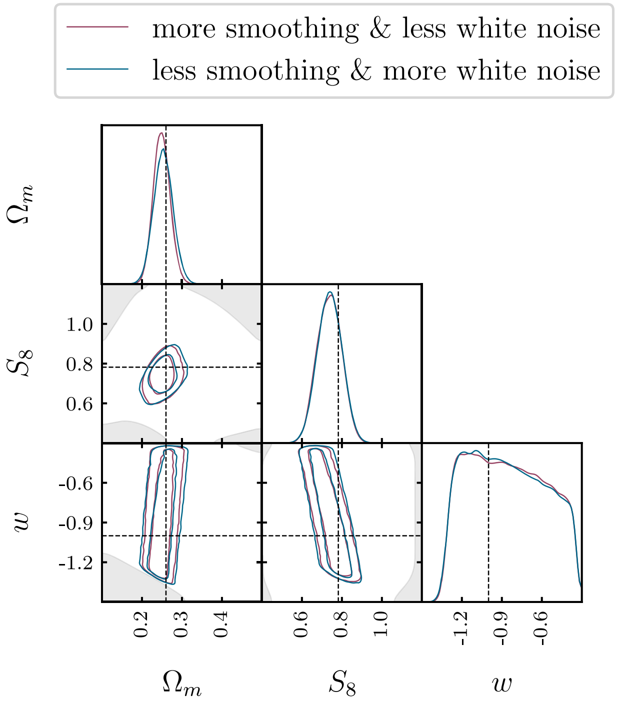

FIG. 18. Comparison of different trade-offs between Gaussian smoothing scale $\sigma_{\min}$ and white-noise level σ_{noise} for fixed multipole $\tilde{\ell}_{\max}$. The marginal contours show no significant differences.

2. Validation

The real-space smoothing procedure detailed in Appendix A 1 allows different trade-offs between the Gaussian smoothing kernel size $\sigma_{\min}$ and pixelwise white-noise amplitude σ_{noise}. The same maximal multipole $\tilde{\ell}_{\max}$ (see Fig. 17) in harmonic space can be achieved by either increasing the smoothing scale while reducing the subsequently added white noise, or vice versa.

We test whether different choices yielding approximately equal multipoles $\tilde{\ell}^i_{\max} \approx [267, 390, 509, 610]$ significantly impact posterior contours. Figure 18 shows results for an example clustering-only analysis evaluated on a fiducial COSMOGRIDV1 mock. The "more smoothing and less white noise" configuration is derived from a scale of $R = 16$ Mpc/h with a 1% threshold in Eq. (A1), yielding

$$\sigma^i_{\min} = [27.6, 18.9, 14.5, 12.1] \text{ arc min},$$
$$\sigma^i_{\text{noise}} = [0.52, 0.21, 0.15, 0.14],$$

while the "less smoothing and more white noise" setup uses a 10% threshold, resulting in

$$\sigma^i_{\min} = [19.5, 13.4, 10.2, 8.5] \text{ arc min},$$
$$\sigma^i_{\text{noise}} = [1.61, 0.67, 0.47, 0.43].$$

The posterior contours show no significant differences.

APPENDIX B: ADDITIONAL FIGURES

(a) Map versus two-point level: We present the full parameter space versions of the posterior results discussed in Sec. VII A in Figs. 19–21. The intrinsic alignment parameters $\eta_{A_{\rm IA}}$, controlling the redshift evolution, and $b_{\rm TA}$, coupling intrinsic alignment to the local density field, are weakly constrained with no significant difference between the map-level and two-point analyses. This is expected since the true $A_{\rm IA} = 0$ for BUZZARD, eliminating any dependence on $\eta_{A_{\rm IA}}$ and $b_{\rm TA}$ according to Eq. (7). The galaxy clustering biases b_g^1–b_g^4 show mild redshift-bin dependence. In all cases, their degeneracy with the S_8 parameter is significantly reduced in the map-level analysis. For the combined probes, adding weak lensing to galaxy clustering further reduces this degeneracy.

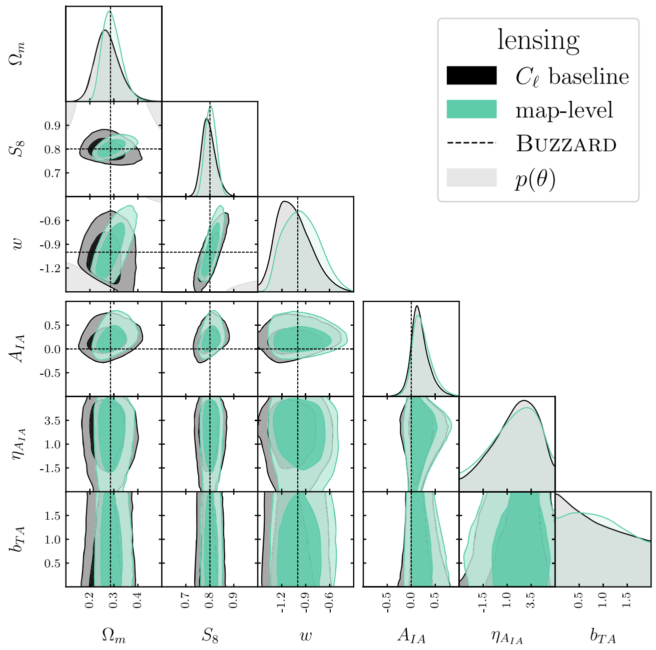

FIG. 19. Like Fig. 11 lensing, but for all parameters.

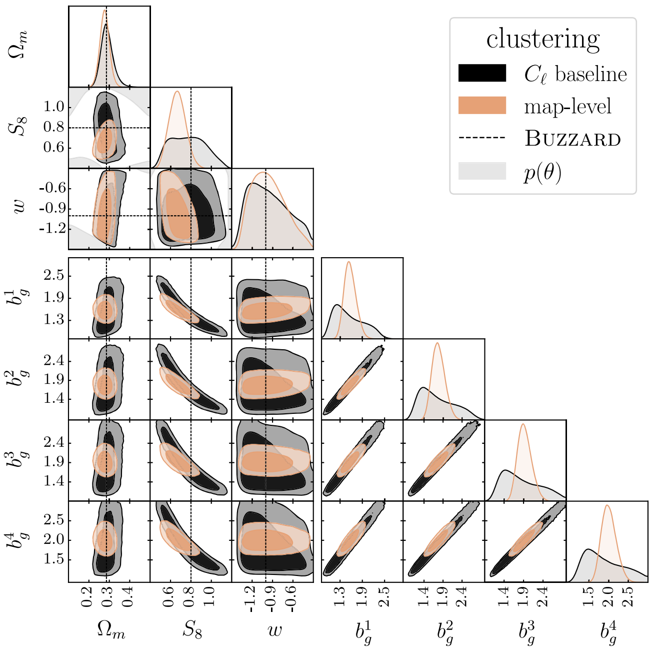

FIG. 20. Like Fig. 11 clustering, but for all parameters.

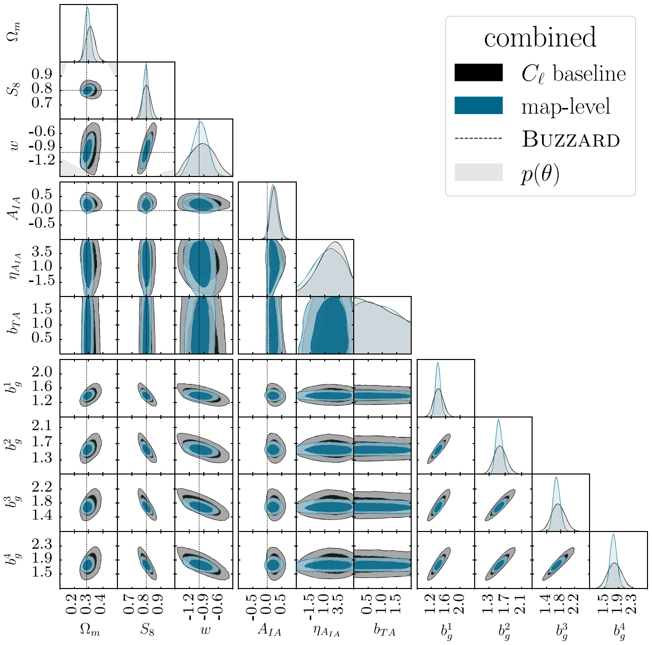

FIG. 21. Like Fig. 12 combined, but for all parameters.

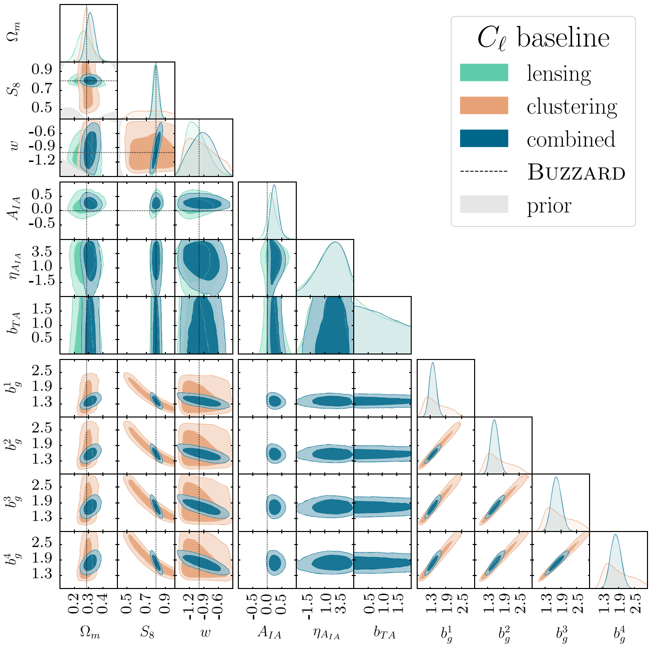

FIG. 22. Like Fig. 14, but for power-spectrum instead of map-level analysis.

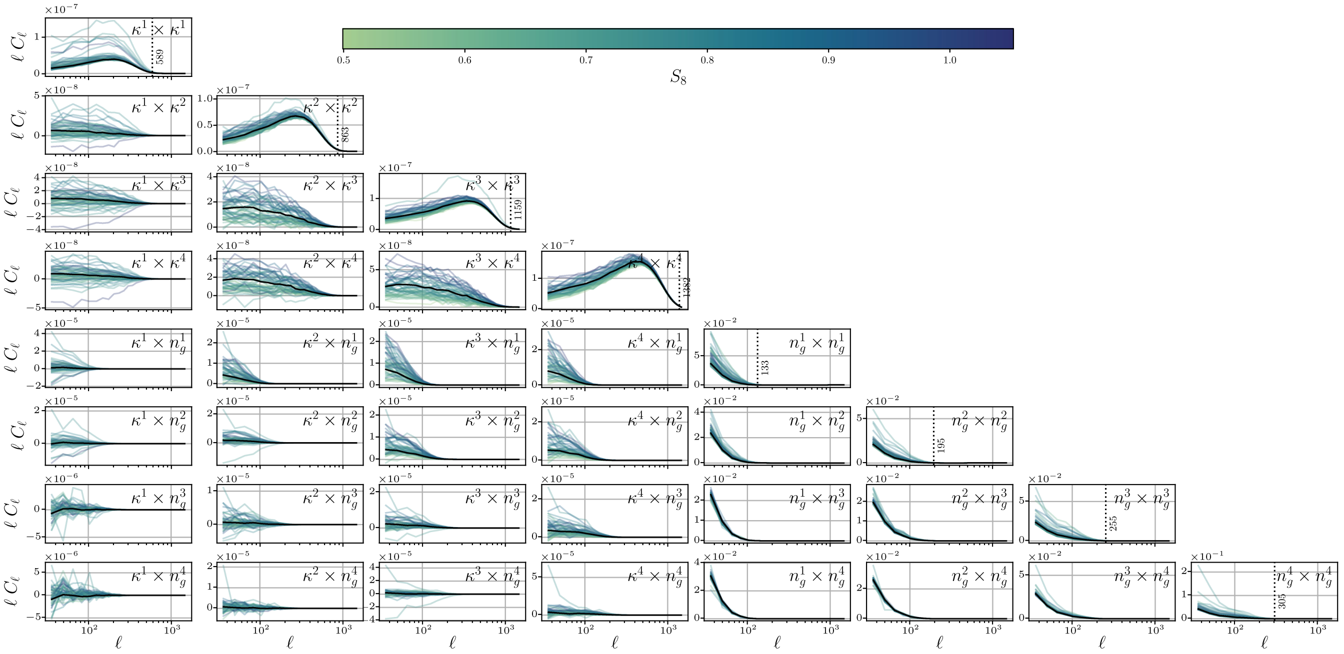

FIG. 23. Auto-, cross redshift bin, and cross-probe power spectra. The colored lines correspond to 50 random COSMOGRIDV1 grid cosmologies, while the black line is obtained from the mean over the 15 BUZZARD realizations.

(b) Probe comparison: Figure 22 compares the different probe configurations for the power spectrum instead of the map-level analysis in Fig. 14.

(c) Power spectrum: In Fig. 23, we show all auto- and cross-power spectra for weak lensing and galaxy clustering used in the analysis. The scale cuts for the power spectra are implemented consistently with the map-level approach described in Sec. VI A, yielding suppression to near zero for scales smaller than the Gaussian smoothing kernel size.

APPENDIX C: INTRINSIC ALIGNMENT

This appendix presents additional results regarding the source galaxy intrinsic alignments modeled according to Sec. III D 3.

1. Projection effects

Due to projection effects, the posterior contours of the A_{IA} parameter conditioned on the mean BUZZARD mock appear to exhibit nontrivial bias in Figs. 11, 12, and 14.

We show the posterior densities for the per-bin values A^i_{IA} computed from Eq. (8) in Fig. 24, where i indexes the four redshift bins of the METACAL source galaxy sample. These two-dimensional marginals demonstrate that the highest posterior probabilities lie close to the true value of zero, consistent with the absence of intrinsic alignment modeling in our BUZZARD mocks (see Sec. V B 1). We therefore conclude that the shift from zero toward positive A_{IA} values in Figs. 11, 12, and 14 results from a projection effect arising from the power-law parametrization of A^i_{IA} in terms of A_{IA} and $\eta_{A_{\mathrm{IA}}}$.

2. COSMOGRIDV1 internal mocks

Since we do not include intrinsic alignments in our BUZZARD mocks, we compare different intrinsic alignment amplitudes A_{IA} using mock observations derived from our weak lensing COSMOGRIDV1 forward model detailed in Sec. III D.

We present results for $A_{\mathrm{IA}} \in \{-1, 0, 1\}$, $\eta_{A_{\mathrm{IA}}} = 0$, and $b_{\mathrm{TA}} = 0$ in Fig. 25 and find that our inference pipeline accurately recovers the ground truth values in all cases. The parameters $\eta_{A_{\mathrm{IA}}}$ and b_{TA} are only weakly constrained and partially hit the prior boundaries. Everything besides A_{IA} being fixed, the marginal distributions of the cosmological parameters are only mildly affected.

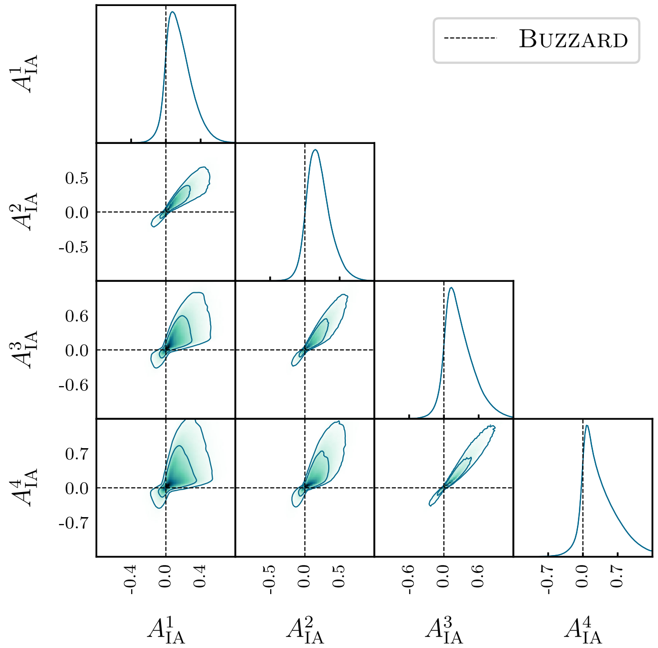

FIG. 24. Marginal posterior probability densities of the per-bin intrinsic alignment amplitude $A^i_{\rm IA}$ parametrized by $A_{\rm IA}$ and $\eta_{A_{\rm IA}}$ according to Eq. (8). The analysis configuration matches Fig. 11(a).

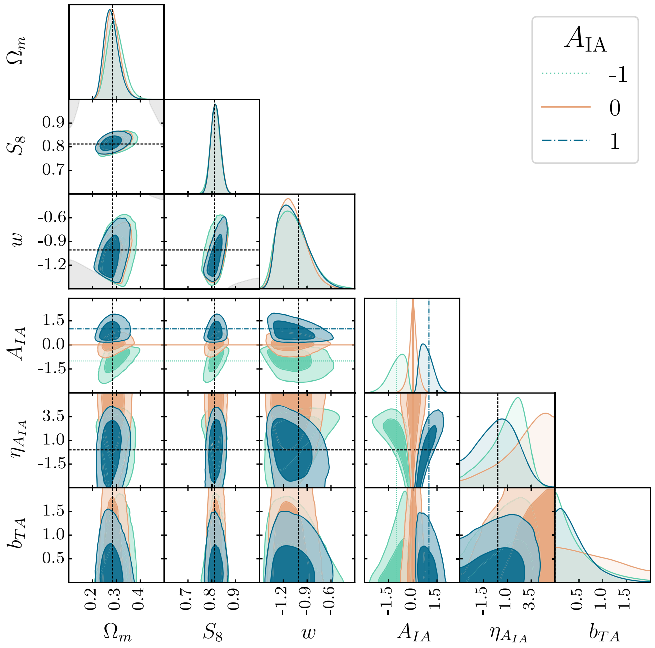

FIG. 25. Marginal posterior probability densities for CosmoGridV1 internal mocks of varying intrinsic alignment amplitude $A_{\rm IA}$. For the intrinsic alignment parameters, the axis limits mark the prior boundaries from Table I. In all cases, the inference pipeline accurately recovers the true parameters.

---

[1] A. Albrecht, G. Bernstein, R. Cahn, W. L. Freedman, J. Hewitt, W. Hu, J. Huth, M. Kamionkowski, E. W. Kolb, L. Knox *et al.*, Report of the dark energy task force, arXiv:astro-ph/0609591v1.

[2] P. Schneider, Weak gravitational lensing, in *Gravitational Lensing: Strong, Weak and Micro*, edited by P. Schneider, C. S. Kochanek, and J. Wambsganss (Springer, Berlin, Heidelberg, 2006), pp. 269–451.

[3] M. Kilbinger, Cosmology with cosmic shear observations: A review, Rep. Prog. Phys. **78**, 086901 (2015).

[4] H. Zhan and J. A. Tyson, Cosmology with the Large Synoptic Survey Telescope: An overview, Rep. Prog. Phys. **81**, 066901 (2018).

[5] T. M. C. Abbott, M. Aguena, A. Alarcon, S. Allam, O. Alves, A. Amon, F. Andrade-Oliveira, J. Annis, S. Avila *et al.* (DES Collaboration), Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing, Phys. Rev. D **105**, 023520 (2022).

[6] C. Heymans, T. Tröster, M. Asgari, C. Blake, H. Hildebrandt, B. Joachimi, K. Kuijken, C.-A. Lin, A. G. Sánchez, J. L. Busch *et al.*, KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints, Astron. Astrophys. **646**, A140 (2021).

[7] E. Krause, X. Fang, S. Pandey, L. F. Secco, O. Alves, H. Huang, J. Blazek, J. Prat, J. Zuntz, T. F. Eifler *et al.*, Dark Energy Survey Year 3 Results: Multi-probe modeling strategy and validation, arXiv:2105.13548.

[8] T. Kacprzak and J. Fluri, DeepLSS: Breaking parameter degeneracies in large-scale structure with deep-learning analysis of combined probes, Phys. Rev. X **12**, 031029 (2022).

[9] https://www.darkenergysurvey.org/.

[10] https://kids.strw.leidenuniv.nl/.

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

[12] https://delve-survey.github.io/.

[13] S. Sugiyama, H. Miyatake, S. More, X. Li, M. Shirasaki, M. Takada, Y. Kobayashi, R. Takahashi, T. Nishimichi, A. J. Nishizawa *et al.*, Hyper Suprime-Cam Year 3 results: Cosmology from galaxy clustering and weak lensing with HSC and SDSS using the minimal bias model, Phys. Rev. D **108**, 123521 (2023).

[14] D. Anbajagane, C. Chang, A. Drlica-Wagner, C. Y. Tan, M. Adamow, R. A. Gruendl, L. F. Secco, Z. Zhang, M. R. Becker, P. S. Ferguson *et al.*, The dark energy camera all data everywhere cosmic shear project V: Constraints on cosmology and astrophysics from 270 million galaxies across 13,000 deg^2 of the sky, arXiv:2509.03582.

[15] https://rubinobservatory.org/explore/lsst.

[16] https://www.euclid-ec.org/public/data/surveys/.

[17] https://science.nasa.gov/mission/roman-space-telescope.

[18] J. A. Peacock, *Cosmological Physics* (Cambridge University Press, Cambridge, England, 1998).

[19] M. Bartelmann and P. Schneider, Weak gravitational lensing, Phys. Rep. **340**, 291 (2001).

[20] A. Amon, D. Gruen, M. A. Troxel, N. MacCrann, S. Dodelson, A. Choi, C. Doux, L. F. Secco, S. Samuroff *et al.* (DES Collaboration), Dark Energy Survey Year 3 results: Cosmology from cosmic shear and robustness to data calibration, Phys. Rev. D **105**, 023514 (2022).

[21] M. Asgari, C.-A. Lin, B. Joachimi, B. Giblin, C. Heymans, H. Hildebrandt, A. Kannawadi, B. Stölzner, T. Tröster, J. L. van den Busch *et al.*, KiDS-1000 cosmology: Cosmic shear constraints and comparison between two point statistics, Astron. Astrophys. **645**, A104 (2021).

[22] C. Doux, B. Jain, D. Zeurcher, J. Lee, X. Fang, R. Rosenfeld, A. Amon, H. Camacho, A. Choi, L. F. Secco *et al.*, Dark Energy Survey Year 3 results: Cosmological constraints from the analysis of cosmic shear in harmonic space, Mon. Not. R. Astron. Soc. **515**, 1942 (2022).

[23] T. Kacprzak, J. Fluri, A. Schneider, A. Refregier, and J. Stadel, CosmoGridV1: A simulated wCDM theory prediction for map-level cosmological inference, J. Cosmol. Astropart. Phys. 02 (2023) 050.

[24] V. Ajani, M. Baldi, A. Barthelemy, A. Boyle, P. Burger, V. F. Cardone, S. Cheng, S. Codis, C. Giocoli *et al.* (Euclid Collaboration), Euclid preparation: XXVIII. Forecasts for ten different higher-order weak lensing statistics, Astron. Astrophys. **675**, A120 (2023).

[25] T. Kacprzak, D. Kirk, O. Friedrich, A. Amara, A. Refregier, L. Marian, J. P. Dietrich, E. Suchyta, J. Aleksić, D. Bacon *et al.*, Cosmology constraints from shear peak statistics in Dark Energy Survey Science Verification data, Mon. Not. R. Astron. Soc. **463**, 3653 (2016).

[26] D. Zürcher, J. Fluri, R. Sgier, T. Kacprzak, M. Gatti, C. Doux, L. Whiteway, A. Réfrégier, C. Chang, N. Jeffrey *et al.*, Dark Energy Survey Year 3 results: Cosmology with peaks using an emulator approach, Mon. Not. R. Astron. Soc. **511**, 2075 (2022).

[27] N. Jeffrey, L. Whiteway, M. Gatti, J. Williamson, J. Alsing, A. Porredon, J. Prat, C. Doux, B. Jain, C. Chang *et al.*, Dark Energy Survey Year 3 results: Likelihood-free, simulation-based wCDM inference with neural compression of weak-lensing map statistics, Mon. Not. R. Astron. Soc. **536**, 1303 (2025).

[28] M. Gatti, C. Chang, O. Friedrich, B. Jain, D. Bacon, M. Crocce, J. DeRose, I. Ferrero, P. Fosalba, E. Gaztanaga *et al.*, Dark Energy Survey Year 3 results: Cosmology with moments of weak lensing mass maps—validation on simulations, Mon. Not. R. Astron. Soc. **498**, 4060 (2020).

[29] M. Gatti, B. Jain, C. Chang, M. Raveri, D. Zürcher, L. Secco, L. Whiteway, N. Jeffrey, C. Doux *et al.* (DES Collaboration), Dark Energy Survey Year 3 results: Cosmology with moments of weak lensing mass maps, Phys. Rev. D **106**, 083509 (2022).

[30] M. Gatti, N. Jeffrey, L. Whiteway, J. Williamson, B. Jain, V. Ajani, D. Anbajagane, G. Giannini, C. Zhou *et al.* (DES Collaboration), Dark Energy Survey Year 3 results: Simulation-based cosmological inference with wavelet harmonics, scattering transforms, and moments of weak lensing mass maps. Validation on simulations, Phys. Rev. D **109**, 063534 (2024).

[31] M. Gatti, G. Campailla, N. Jeffrey, L. Whiteway, A. Porredon, J. Prat, J. Williamson, M. Raveri, B. Jain *et al.* (Dark Energy Survey Collaboration), Dark Energy Survey Year 3 results: Simulation-based cosmological inference with wavelet harmonics, scattering transforms, and moments of weak lensing mass maps. II. cosmological results, Phys. Rev. D **111**, 063504 (2025).

[32] L. F. Secco, M. Jarvis, B. Jain, C. Chang, M. Gatti, J. Frieman, S. Adhikari, A. Alarcon, A. Amon *et al.* (DES Collaboration), Dark Energy Survey Year 3 Results: Three-point shear correlations and mass aperture moments, Phys. Rev. D **105**, 103537 (2022).

[33] R. C. H. Gomes, S. Sugiyama, B. Jain, M. Jarvis, D. Anbajagane, M. Gatti, D. Gebauer, Z. Gong, A. Halder, G. A. Marques *et al.*, Cosmology with second and third-order shear statistics for the Dark Energy Survey: Methods and simulated analysis, Phys. Rev. D **112**, 123514 (2025).

[34] R. C. H. Gomes, S. Sugiyama, B. Jain, M. Jarvis, D. Anbajagane, A. Halder, G. A. Marques, S. Pandey, J. Marshall, A. Alarcon *et al.*, Dark Energy Survey Year 3 Results: Cosmological constraints from second and third-order shear statistics, Phys. Rev. D **112**, 123515 (2025).

[35] D. Anbajagane, C. Chang, A. Banerjee, T. Abel, M. Gatti, V. Ajani, A. Alarcon, A. Amon, E. J. Baxter, K. Bechtol *et al.*, Beyond the 3rd moment: A practical study of using lensing convergence CDFs for cosmology with DES Y3, Mon. Not. R. Astron. Soc. **526**, 5530 (2023).

[36] J. Prat, M. Gatti, C. Doux, P. Pranav, C. Chang, N. Jeffrey, L. Whiteway, D. Anbajagane, S. Sugiyama, A. Thomsen *et al.*, Dark Energy Survey Year 3 results: wCDM cosmology from simulation-based inference with persistent homology on the sphere, Mon. Not. R. Astron. Soc. **545**, staf2152 (2026).

[37] J. Fluri, T. Kacprzak, A. Refregier, A. Amara, A. Lucchi, and T. Hofmann, Cosmological constraints from noisy convergence maps through deep learning, Phys. Rev. D **98**, 123518 (2018).

[38] J. Fluri, T. Kacprzak, A. Lucchi, A. Refregier, A. Amara, T. Hofmann, and A. Schneider, Cosmological constraints with deep learning from KiDS-450 weak lensing maps, Phys. Rev. D **100**, 063514 (2019).

[39] J. Fluri, T. Kacprzak, A. Refregier, A. Lucchi, and T. Hofmann, Cosmological parameter estimation and inference using deep summaries, Phys. Rev. D **104**, 123526 (2021).

[40] J. Fluri, T. Kacprzak, A. Lucchi, A. Schneider, A. Refregier, and T. Hofmann, Full ωCDM analysis of KiDS-1000 weak lensing maps using deep learning, Phys. Rev. D **105**, 083518 (2022).

[41] N. Jeffrey, J. Alsing, and F. Lanusse, Likelihood-free inference with neural compression of DES SV weak lensing map statistics, Mon. Not. R. Astron. Soc. **501**, 954 (2021).

[42] A. Gupta, J. M. Z. Matilla, D. Hsu, and Z. Haiman, Non-Gaussian information from weak lensing data via deep learning, Phys. Rev. D **97**, 103515 (2018).

[43] D. Ribli, B. Á. Pataki, J. M. Zorrilla Matilla, D. Hsu, Z. Haiman, and I. Csabai, Weak lensing cosmology with

convolutional neural networks on noisy data, Mon. Not. R. Astron. Soc. **490**, 1843 (2019).
[44] J. M. Z. Matilla, M. Sharma, D. Hsu, and Z. Haiman, Interpreting deep learning models for weak lensing, Phys. Rev. D **102**, 123506 (2020).
[45] T. Lu, Z. Haiman, and J. M. Zorrilla Matilla, Simultaneously constraining cosmology and baryonic physics via deep learning from weak lensing, Mon. Not. R. Astron. Soc. **511**, 1518 (2022).
[46] T. Lu, Z. Haiman, and X. Li, Cosmological constraints from HSC survey first-year data using deep learning, Mon. Not. R. Astron. Soc. **521**, 2050 (2023).
[47] A. Akhmetzhanova, S. Mishra-Sharma, and C. Dvorkin, Data compression and inference in cosmology with self-supervised machine learning, Mon. Not. R. Astron. Soc. **527**, 7459 (2024).
[48] D. Lanzieri, J. Zeghal, T. L. Makinen, A. Boucaud, J.-L. Starck, and F. Lanusse, Optimal neural summarization for full-field weak lensing cosmological implicit inference, Astron. Astrophys. **697**, A162 (2025).
[49] K. Cranmer, J. Brehmer, and G. Louppe, The frontier of simulation-based inference, Proc. Natl. Acad. Sci. U.S.A. **117**, 30055 (2020).
[50] G. Papamakarios, Neural density estimation and likelihood-free inference, arXiv:1910.13233.
[51] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, Normalizing flows for probabilistic modeling and inference, J. Mach. Learn. Res. **22**, 57 (2021).
[52] J. Alsing, T. Charnock, S. Feeney, and B. Wandelt, Fast likelihood-free cosmology with neural density estimators and active learning, Mon. Not. R. Astron. Soc. **488**, 4440 (2019).
[53] G. Papamakarios, D. Sterratt, and I. Murray, Sequential neural likelihood: Fast likelihood-free inference with autoregressive flows, in *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics* (PMLR, 2019), pp. 837–848.
[54] J. Alsing, B. Wandelt, and S. Feeney, Massive optimal data compression and density estimation for scalable, likelihood-free inference in cosmology, Mon. Not. R. Astron. Soc. **477**, 2874 (2018).
[55] J. DeRose, R. H. Wechsler, M. R. Becker, E. S. Rykoff, S. Pandey, N. MacCrann, A. Amon, J. Myles, E. Krause *et al.* (DES Collaboration), Dark Energy Survey Year 3 results: Cosmology from combined galaxy clustering and lensing validation on cosmological simulations, Phys. Rev. D **105**, 123520 (2022).
[56] The Dark Energy Survey Collaboration, The Dark Energy Survey, arXiv:astro-ph/0510346.
[57] T. Abbott, F. B. Abdalla, J. Aleksić, S. Allam, A. Amara, D. Bacon, E. Balbinot, M. Banerji, K. Bechtol *et al.* (Dark Energy Survey Collaboration), The Dark Energy Survey: More than dark energy—an overview, Mon. Not. R. Astron. Soc. **460**, 1270 (2016).
[58] B. Flaugher, H. T. Diehl, K. Honscheid, T. M. C. Abbott, O. Alvarez, R. Angstadt, J. T. Annis, M. Antonik, O. Ballester, L. Beaufore *et al.*, The dark energy camera, Astron. J. **150**, 150 (2015).
[59] E. Morganson, R. A. Gruendl, F. Menanteau, M. Carrasco Kind, Y. C. Chen, G. Daues, A. Drlica-Wagner, D. N. Friedel, M. Gower, M. W. G. Johnson *et al.*, The Dark Energy Survey image processing pipeline, Publ. Astron. Soc. Pac. **130**, 074501 (2018).
[60] T. M. C. Abbott, F. B. Abdalla, S. Allam, A. Amara, J. Annis, J. Asorey, S. Avila, O. Ballester, M. Banerji, W. Barkhouse *et al.*, The Dark Energy Survey: Data release 1, Astrophys. J. Suppl. Ser. **239**, 18 (2018).
[61] M. Gatti, E. Sheldon, A. Amon, M. Becker, M. Troxel, A. Choi, C. Doux, N. MacCrann, A. Navarro-Alsina, I. Harrison *et al.*, Dark Energy Survey Year 3 results: Weak lensing shape catalogue, Mon. Not. R. Astron. Soc. **504**, 4312 (2021).
[62] E. Huff and R. Mandelbaum, Metacalibration: Direct self-calibration of biases in shear measurement, arXiv:1702.02600.
[63] E. S. Sheldon and E. M. Huff, Practical weak-lensing shear measurement with metacalibration, Astrophys. J. **841**, 24 (2017).
[64] N. MacCrann, M. R. Becker, J. McCullough, A. Amon, D. Gruen, M. Jarvis, A. Choi, M. A. Troxel, E. Sheldon, B. Yanny *et al.*, Dark Energy Survey Y3 results: Blending shear and redshift biases in image simulations, Mon. Not. R. Astron. Soc. **509**, 3371 (2022).
[65] J. Myles, A. Alarcon, A. Amon, C. Sánchez, S. Everett, J. DeRose, J. McCullough, D. Gruen, G. M. Bernstein, M. A. Troxel *et al.*, Dark Energy Survey Year 3 results: Redshift calibration of the weak lensing source galaxies, Mon. Not. R. Astron. Soc. **505**, 4249 (2021).
[66] M. Gatti, G. Giannini, G. M. Bernstein, A. Alarcon, J. Myles, A. Amon, R. Cawthon, M. Troxel, J. DeRose, S. Everett *et al.*, Dark Energy Survey Year 3 Results: Clustering redshifts—calibration of the weak lensing source redshift distributions with redMaGiC and BOSS/eBOSS, Mon. Not. R. Astron. Soc. **510**, 1223 (2022).
[67] A. Porredon, M. Crocce, P. Fosalba, J. Elvin-Poole, A. Carnero Rosell, R. Cawthon, T. F. Eifler, X. Fang, I. Ferrero, E. Krause *et al.*, Dark Energy Survey Year 3 results: Optimizing the lens sample in a combined galaxy clustering and galaxy-galaxy lensing analysis, Phys. Rev. D **103**, 043503 (2021).
[68] A. Porredon, M. Crocce, J. Elvin-Poole, R. Cawthon, G. Giannini, J. De Vicente, A. Carnero Rosell, I. Ferrero, E. Krause *et al.* (DES Collaboration), Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and galaxy-galaxy lensing using the MagLim lens sample, Phys. Rev. D **106**, 103530 (2022).
[69] I. Sevilla-Noarbe, K. Bechtol, M. C. Kind, A. C. Rosell, M. R. Becker, A. Drlica-Wagner, R. A. Gruendl, E. S. Rykoff, E. Sheldon, B. Yanny *et al.*, Dark Energy Survey Year 3 results: Photometric data set for cosmology, Astrophys. J. Suppl. Ser. **254**, 24 (2021).
[70] J. De Vicente, E. Sánchez, and I. Sevilla-Noarbe, DNF—Galaxy photometric redshift by Directional Neighbourhood Fitting, Mon. Not. R. Astron. Soc. **459**, 3078 (2016).
[71] C.-A. Lin and M. Kilbinger, A new model to predict weak-lensing peak counts—I. Comparison with N-body simulations, Astron. Astrophys. **576**, A24 (2015).

[72] N. Porqueres, A. Heavens, D. Mortlock, and G. Lavaux, Bayesian forward modelling of cosmic shear data, Mon. Not. R. Astron. Soc. **502**, 3035 (2021).

[73] M. Von Wietersheim-Kramsta, K. Lin, N. Tessore, B. Joachimi, A. Loureiro, R. Reischke, and A. H. Wright, KiDS-SBI: Simulation-based inference analysis of KiDS-1000 cosmic shear, Astron. Astrophys. **694**, A223 (2025).

[74] http://www.cosmogrid.ai/.

[75] A. Thomsen, Multiprobe-simulation-forward-model (2026), https://github.com/des-science/multiprobe-simulation-forward-model.

[76] D. Potter, J. Stadel, and R. Teyssier, PKDGRAV3: Beyond trillion particle cosmological simulations for the next era of galaxy surveys, Comput. Astrophys. Cosmol. **4**, 2 (2017).

[77] T. M. C. Abbott, M. Acevedo, M. Adamow, M. Aguena, A. Alarcon, S. Allam, O. Alves, F. Andrade-Oliveira, J. Annis *et al.* (DES Collaboration), Dark Energy Survey: Implications for cosmological expansion models from the final DES Baryon Acoustic Oscillation and Supernova data, arXiv:2503.06712.

[78] A. Adame, J. Aguilar, S. Ahlen, S. Alam, D. Alexander, M. Alvarez, O. Alves, A. Anand, U. Andrade, E. Armengaud *et al.*, DESI 2024 VI: Cosmological constraints from the measurements of baryon acoustic oscillations, J. Cosmol. Astropart. Phys. 02 (2025) 021.

[79] A. Adame, J. Aguilar, S. Ahlen, S. Alam, D. Alexander, C. Allende Prieto, M. Alvarez, O. Alves, A. Anand, U. Andrade *et al.*, DESI 2024 VII: Cosmological constraints from the full-shape modeling of clustering measurements, J. Cosmol. Astropart. Phys. 07 (2025) 028.

[80] M. Abdul Karim, J. Aguilar, S. Ahlen, S. Alam, L. Allen, C. A. Prieto, O. Alves, A. Anand, U. Andrade *et al.* (DESI Collaboration), DESI DR2 results. II. Measurements of baryon acoustic oscillations and cosmological constraints, Phys. Rev. D **112**, 083515 (2025).

[81] N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak *et al.*, Planck 2018 results—VI. Cosmological parameters, Astron. Astrophys. **641**, A6 (2020).

[82] I. M. Sobol', On the distribution of points in a cube and the approximate evaluation of integrals, USSR Computational Mathematics and Mathematical Physics **7**, 86 (1967).

[83] https://healpix.sourceforge.io/.

[84] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, HEALPIX: A framework for high-resolution discretization and fast analysis of data distributed on the sphere, Astrophys. J. **622**, 759 (2005).

[85] M. P. van Daalen, J. Schaye, C. M. Booth, and C. Dalla Vecchia, The effects of galaxy formation on the matter power spectrum: A challenge for precision cosmology, Mon. Not. R. Astron. Soc. **415**, 3649 (2011).

[86] I. G. McCarthy, J. Schaye, S. Bird, and A. M. C. Le Brun, The bahamas project: Calibrated hydrodynamical simulations for large-scale structure cosmology, Mon. Not. R. Astron. Soc. **465**, 2936 (2017).

[87] E. Chisari, A. J. Mead, S. Joudaki, P. G. Ferreira, A. Schneider, J. Mohr, T. Tröster, D. Alonso, I. G. McCarthy, S. Martin-Alvarez *et al.*, Modelling baryonic feedback for survey cosmology, Open J. Astrophys. **2** (2019), 10.21105/astro.1905.06082.

[88] F. Villaescusa-Navarro, D. Anglés-Alcázar, S. Genel, D. N. Spergel, R. S. Somerville, R. Dave, A. Pillepich, L. Hernquist, D. Nelson, P. Torrey *et al.*, The CAMELS Project: Cosmology and astrophysics with machine-learning simulations, Astrophys. J. **915**, 71 (2021).

[89] J. Schaye, C. D. Vecchia, C. M. Booth, R. P. C. Wiersma, T. Theuns, M. R. Haas, S. Bertone, A. R. Duffy, I. G. McCarthy, and F. van de Voort, The physics driving the cosmic star formation history, Mon. Not. R. Astron. Soc. **402**, 1536 (2010).

[90] A. Schneider and R. Teyssier, A new method to quantify the effects of baryons on the matter power spectrum, J. Cosmol. Astropart. Phys. 12 (2015) 049.

[91] Schneider, R. Teyssier, J. Stadel, N. E. Chisari, A. M. C. L. Brun, A. Amara, and A. Refregier, Quantifying baryon effects on the matter power spectrum and the weak lensing shear correlation, J. Cosmol. Astropart. Phys. 03 (2019) 020.

[92] A. Schneider, S. K. Giri, S. Amodeo, and A. Refregier, Constraining baryonic feedback and cosmology with weak-lensing, X-ray, and kinematic Sunyaev–Zeldovich observations, Mon. Not. R. Astron. Soc. **514**, 3802 (2022).

[93] S. K. Giri and A. Schneider, Emulation of baryonic effects on the matter power spectrum and constraints from galaxy cluster data, J. Cosmol. Astropart. Phys. 12 (2021) 046.

[94] https://cosmology.ethz.ch/research/software-lab/UFalcon.html.

[95] R. J. Sgier, A. Réfrégier, A. Amara, and A. Nicola, Fast generation of covariance matrices for weak lensing, J. Cosmol. Astropart. Phys. 01 (2019) 044.

[96] R. Sgier, J. Fluri, J. Herbel, A. Réfrégier, A. Amara, T. Kacprzak, and A. Nicola, Fast lightcones for combined cosmological probes, J. Cosmol. Astropart. Phys. 02 (2021) 047.

[97] A. Reeves, A. Nicola, A. Refregier, T. Kacprzak, and L. F. M. P. Valle, 12×2 pt combined probes: Pipeline, neutrino mass, and data compression, J. Cosmol. Astropart. Phys. 01 (2024) 042.

[98] D. Zürcher, J. Fluri, R. Sgier, T. Kacprzak, and A. Refregier, Cosmological forecast for non-Gaussian statistics in large-scale weak lensing surveys, J. Cosmol. Astropart. Phys. 01 (2021) 028.

[99] D. Zürcher, J. Fluri, V. Ajani, S. Fischbacher, A. Refregier, and T. Kacprzak, Towards a full wCDM map-based analysis for weak lensing surveys, Mon. Not. R. Astron. Soc. **525**, 761 (2023).

[100] A. Reeves, A. Nicola, and A. Refregier, Tuning the cosmic instrument: Robust cosmology through combined probes, J. Cosmol. Astropart. Phys. 06 (2025) 042.

[101] A. Reeves, S. Ferraro, A. Nicola, and A. Refregier, Multiprobe constraints on early and late time dark energy, arXiv:2510.06114.

[102] A. Petri, Z. Haiman, and M. May, Validity of the Born approximation for beyond Gaussian weak lensing observables, Phys. Rev. D **95**, 123503 (2017).

[103] P. Fosalba, E. Gaztañaga, F. J. Castander, and M. Manera, The onion universe: All sky lightcone simulations in

spherical shells, Mon. Not. R. Astron. Soc. **391**, 435 (2008).

[104] S. Bridle and L. King, Dark energy constraints from cosmic shear power spectra: Impact of intrinsic alignments on photometric redshift requirements, New J. Phys. **9**, 444 (2007).

[105] N. Jeffrey and B. D. Wandelt, Solving high-dimensional parameter inference: Marginal posterior densities & Moment Networks, arXiv:2011.05991.

[106] https://github.com/LSSTDESC/CCL.

[107] N. E. Chisari, D. Alonso, E. Krause, C. D. Leonard, P. Bull, J. Neveu, A. Villarreal, S. Singh, T. McClintock, J. Ellison *et al.*, Core cosmology library: Precision cosmological predictions for LSST, Astrophys. J. Suppl. Ser. **242**, 2 (2019).

[108] https://cosmology.ethz.ch/research/software-lab/PyCosmo.html.

[109] A. Refregier, L. Gamper, A. Amara, and L. Heisenberg, PYCOSMO: An integrated cosmological Boltzmann solver, Astron. Comput. **25**, 38 (2018).

[110] M. Knabenhans, J. Stadel, D. Potter, J. Dakin, S. Hannestad, T. Tram, S. Marelli, A. Schneider, R. Teyssier *et al.* (Euclid Collaboration), Euclid preparation: IX. EuclidEmulator2—power spectrum emulation with massive neutrinos and self-consistent dark energy perturbations, Mon. Not. R. Astron. Soc. **505**, 2840 (2021).

[111] T. Tan, D. Zürcher, J. Fluri, A. Refregier, F. Tarsitano, and T. Kacprzak, Assessing theoretical uncertainties for cosmological constraints from weak lensing surveys, Mon. Not. R. Astron. Soc. **522**, 3766 (2023).

[112] N. Perraudin, M. Defferrard, T. Kacprzak, and R. Sgier, DEEPSPHERE: Efficient spherical convolutional neural network with HEALPIX sampling for cosmological applications, Astron. Comput. **27**, 130 (2019).

[113] M. Defferrard, M. Milani, F. Gusset, and N. Perraudin, DEEPSPHERE: A graph-based spherical CNN, in *International Conference on Learning Representations* (2019).

[114] N. Jeffrey, M. Gatti, C. Chang, L. Whiteway, U. Demirbozan, A. Kovacs, G. Pollina, D. Bacon, N. Hamaus, T. Kacprzak *et al.*, Dark Energy Survey Year 3 results: Curved-sky weak lensing mass map reconstruction, Mon. Not. R. Astron. Soc. **505**, 4626 (2021).

[115] N. Kaiser and G. Squires, Mapping the dark matter with weak gravitational lensing, Astrophys. J. **404**, 441 (1993).

[116] C. G. R. Wallis, M. A. Price, J. D. McEwen, T. D. Kitching, B. Leistedt, and A. Plouviez, Mapping dark matter on the celestial sphere with weak gravitational lensing, Mon. Not. R. Astron. Soc. **509**, 4480 (2022).

[117] M. Gatti, N. Jeffrey, L. Whiteway, V. Ajani, T. Kacprzak, D. Zürcher, C. Chang, B. Jain, J. Blazek, E. Krause *et al.*, Detection of the significant impact of source clustering on higher order statistics with DES Year 3 weak gravitational lensing data, Mon. Not. R. Astron. Soc. **527**, L115 (2023).

[118] M. A. Troxel and M. Ishak, The intrinsic alignment of galaxies and its impact on weak gravitational lensing in an era of precision cosmology, Phys. Rep. **558**, 1 (2015).

[119] J. A. Blazek, N. MacCrann, M. A. Troxel, and X. Fang, Beyond linear galaxy alignments, Phys. Rev. D **100**, 103506 (2019).

[120] J. Harnois-Déraps, N. Martinet, and R. Reischke, Cosmic shear beyond 2-point statistics: Accounting for galaxy intrinsic alignment with projected tidal fields, Mon. Not. R. Astron. Soc. **509**, 3868 (2021).

[121] Q. Hang, N. Jeffrey, L. Whiteway, O. Lahav, J. Williamson, M. Gatti, J. DeRose, A. Kovacs, A. Alarcon, A. Amon *et al.*, Biasing from galaxy trough and peak profiles with the DES Y3 redMaGiC galaxies and the weak lensing mass map, Mon. Not. R. Astron. Soc. **546**, stag006 (2026).

[122] T. Bayes and n. Price, LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S, Phil. Trans. R. Soc. London **53**, 370 (1763).

[123] A. Hall and A. Taylor, Non-Gaussian likelihood of weak lensing power spectra, Phys. Rev. D **105**, 123527 (2022).

[124] V. Oehl and T. Tröster, The exact non-Gaussian weak lensing likelihood: A framework to calculate analytic likelihoods for correlation functions on masked Gaussian random fields, Open J. Astrophys. **8** (2025), 10.33232/001c.144235.

[125] C.-H. Lin, J. Harnois-Déraps, T. Eifler, T. Pospisil, R. Mandelbaum, A. B. Lee, and S. Singh (The LSST Dark Energy Science Collaboration), Non-Gaussianity in the weak lensing correlation function likelihood—implications for cosmological parameter biases, Mon. Not. R. Astron. Soc. **499**, 2977 (2020).

[126] E. Sellentin, C. Heymans, and J. Harnois-Déraps, The skewed weak lensing likelihood: Why biases arise, despite data and theory being sound, Mon. Not. R. Astron. Soc. **477**, 4879 (2018).

[127] E. Sellentin and A. F. Heavens, On the insufficiency of arbitrarily precise covariance matrices: Non-Gaussian weak-lensing likelihoods, Mon. Not. R. Astron. Soc. **473**, 2355 (2018).

[128] P. L. Taylor, T. D. Kitching, J. Alsing, B. D. Wandelt, S. M. Feeney, and J. D. McEwen, Cosmic shear: Inference from forward models, Phys. Rev. D **100**, 023519 (2019).

[129] J. M. Bernardo and A. F. M. Smith, Bayesian theory, Meas. Sci. Technol. **12**, 221 (2001).

[130] A. Thomsen (2026), https://github.com/des-science/y3-deep-lss.

[131] S. Kullback and R. A. Leibler, On information and sufficiency, Ann. Math. Stat. **22**, 79 (1951).

[132] C. E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. **27**, 379 (1948).

[133] Y. Chen, D. Zhang, M. U. Gutmann, A. Courville, and Z. Zhu, Neural Approximate Sufficient Statistics for Implicit Models, in *International Conference on Learning Representations* (2020).

[134] D. Barber and F. Agakov, The IM algorithm: A variational approach to information maximization, in *Proceedings of the 17th International Conference on Neural Information Processing Systems*, NIPS'03 (MIT Press, Cambridge, MA, USA, 2003), pp. 201–208.

[135] G. Papamakarios and I. Murray, Fast ε-free inference of simulation models with Bayesian conditional density estimation, in *Proceedings of the 30th International Conference on Neural Information Processing Systems*,

NIPS'16 (Curran Associates Inc., Red Hook, New York, USA, 2016), pp. 1036–1044.
[136] J.-M. Lueckmann, P. J. Goncalves, G. Bassetto, K. Öcal, M. Nonnenmacher, and J. H. Macke, Flexible statistical inference for mechanistic models of neural dynamics, arXiv:1711.01861.
[137] D. S. Greenberg, M. Nonnenmacher, and J. H. Macke, Automatic posterior transformation for likelihood-free inference, arXiv:1905.07488.
[138] J.-M. Lueckmann, G. Bassetto, T. Karaletsos, and J. H. Macke, Likelihood-free inference with emulator networks, arXiv:1805.09294.
[139] https://www.tensorflow.org/.
[140] https://github.com/deepsphere, DEEPSPHERE.
[141] J. L. Ba, J. R. Kiros, and G. E. Hinton, Layer normalization, arXiv:1607.06450.
[142] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. **15**, 1929 (2014).
[143] https://docs.nersc.gov/systems/perlmutter/architecture/.
[144] https://www.nersc.gov/.
[145] https://horovod.ai/.
[146] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980.
[147] A. Zonca, L. P. Singer, D. Lenz, M. Reinecke, C. Rosset, E. Hivon, and K. M. Gorski, HEALPY: Equal area pixelization and spherical harmonics transforms for data on the sphere in PYTHON, J. Open Source Software **4**, 1298 (2019).
[148] L. Faga, F. Andrade-Oliveira, H. Camacho, R. Rosenfeld, M. Lima, C. Doux, X. Fang, J. Prat, A. Porredon, M. Aguena *et al.*, Dark Energy Survey Year 3 results: Cosmology from galaxy clustering and galaxy–galaxy lensing in harmonic space, Mon. Not. R. Astron. Soc. **536**, 1586 (2025).
[149] A. Thomsen (2026), https://github.com/des-science/multiprobe-simulation-inference.
[150] F. Arend Torres, M. M. Negri, and J. Aellen, FLOWCONDUCTOR: (Conditional) Normalizing Flows and bijective Layers for PYTORCH, Zenodo, 10.5281/zenodo.13952362 (2024).
[151] M. M. Negri, F. A. Torres, and V. Roth, Conditional matrix flows for Gaussian graphical models, arXiv:2306.07255.
[152] G. Papamakarios, T. Pavlakou, and I. Murray, Masked autoregressive flow for density estimation, in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2017), Vol. 30.
[153] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, EMCEE: The MCMC Hammer, Publ. Astron. Soc. Pac. **125**, 306 (2013).
[154] A. E. Bayer, F. Villaescusa-Navarro, S. Sharief, R. Teyssier, L. H. Garrison, L. Perreault-Levasseur, G. L. Bryan, M. Gatti, and E. Visbal, Field-level comparison and robustness analysis of cosmological N-body simulations, Astrophys. J. **989**, 207 (2025).
[155] J. DeRose, R. H. Wechsler, M. R. Becker, M. T. Busha, E. S. Rykoff, N. MacCrann, B. Erickson, A. E. Evrard, A. Kravtsov, D. Gruen *et al.*, The Buzzard Flock: Dark Energy Survey synthetic sky catalogs, arXiv:1901.02401.
[156] R. H. Wechsler, J. DeRose, M. T. Busha, M. R. Becker, E. Rykoff, and A. Evrard, ADDGALS: Simulated sky catalogs for wide field galaxy surveys, Astrophys. J. **931**, 145 (2022).
[157] P. Berner, Forward modelling of galaxy clustering using subhalo abundance matching, Ph.D. thesis, ETH Zurich, 2024.
[158] M. Asgari, A. J. Mead, and C. Heymans, The halo model for cosmology: A pedagogical review, Open J. Astrophys. **6** (2023), 10.21105/astro.2303.08752.
[159] M. R. Becker, Calclens: Weak lensing simulations for large-area sky surveys and second-order effects in cosmic shear power spectra, Mon. Not. R. Astron. Soc. **435**, 115 (2013).
[160] N. MacCrann, J. DeRose, R. H. Wechsler, J. Blazek, E. Gaztanaga, M. Crocce, E. S. Rykoff, M. R. Becker, B. Jain, E. Krause *et al.*, DES Y1 Results: Validating cosmological parameter estimation using simulated dark energy surveys, Mon. Not. R. Astron. Soc. **480**, 4614 (2018).
[161] S. Everett, B. Yanny, N. Kuropatkin, E. M. Huff, Y. Zhang, J. Myles, A. Masegian, J. Elvin-Poole, S. Allam, G. M. Bernstein *et al.*, Dark Energy Survey Year 3 results: Measuring the survey transfer function with Balrog, Astrophys. J. Suppl. Ser. **258**, 15 (2022).
[162] J. Bucko and DES (to be published).
[163] L. F. Secco, S. Samuroff, E. Krause, B. Jain, J. Blazek, M. Raveri, A. Campos, A. Amon, A. Chen *et al.* (DES Collaboration), Dark Energy Survey Year 3 results: Cosmology from cosmic shear and robustness to modeling uncertainty, Phys. Rev. D **105**, 023515 (2022).
[164] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis Third Edition* (CRC Press, 2013).
[165] J. Hermans, A. Delaunoy, F. Rozet, A. Wehenkel, V. Begy, and G. Louppe, A trust crisis in simulation-based inference? Your posterior approximations can be unfaithful, arXiv:2110.06581.
[166] P. Lemos, A. Coogan, Y. Hezaveh, and L. Perreault-Levasseur, Sampling-based accuracy testing of posterior estimators for general inference, in *Proceedings of the 40th International Conference on Machine Learning* (PMLR, 2023), pp. 19256–19273.
[167] https://github.com/Ciela-Institute/tarp.
[168] K. S. Dawson, D. J. Schlegel, C. P. Ahn, S. F. Anderson, É. Aubourg, S. Bailey, R. H. Barkhouser, J. E. Bautista, A. Beifiori, A. A. Berlind *et al.*, The baryon oscillation spectroscopic survey of SDSS-III, Astron. J. **145**, 10 (2012).
[169] A. Dey, D. J. Schlegel, D. Lang, R. Blum, K. Burleigh, X. Fan, J. R. Findlay, D. Finkbeiner, D. Herrera, S. Juneau *et al.*, Overview of the DESI legacy imaging surveys, Astron. J. **157**, 168 (2019).
[170] O. H. E. Philcox and M. M. Ivanov, BOSS DR12 full-shape cosmology: ΛCDM constraints from the large-scale galaxy power spectrum and bispectrum monopole, Phys. Rev. D **105**, 043517 (2022).
[171] O. H. E. Philcox, M. M. Ivanov, G. Cabass, M. Simonović, M. Zaldarriaga, and T. Nishimichi, Cosmology with the redshift-space galaxy bispectrum monopole at one-loop order, Phys. Rev. D **106**, 043530 (2022).

[172] A. Chudaykin, M. M. Ivanov, and O. H. E. Philcox, Reanalyzing DESI DR1: 1. ΛCDM constraints from the power spectrum and bispectrum, arXiv:2507.13433.

[173] A. J. Zhou, M. Gatti, D. Anbajagane, S. Dodelson, M. Schaller, and J. Schaye, Map-level baryonification: Unified treatment of weak lensing two-point and higher-order statistics, J. Cosmol. Astropart. Phys. 09 (2025) 073.

[174] P. Berner, A. Refregier, B. Moser, L. Tortorelli, L. F. Machado Poletti Valle, and T. Kacprzak, Fast forward modelling of galaxy spatial and statistical distributions, J. Cosmol. Astropart. Phys. 04 (2024) 023.

[175] S. Fischbacher, T. Kacprzak, L. F. Machado Poletti Valle, and A. Refregier, SHAM-OT: Rapid subhalo abundance matching with optimal transport, Mon. Not. R. Astron. Soc.: Lett. **542**, L53 (2025).

[176] S. Fischbacher, T. Kacprzak, L. Tortorelli, B. Moser, A. Refregier, P. Gebhardt, and D. Gruen, GalSBI: Phenomenological galaxy population model for cosmology using simulation-based inference, J. Cosmol. Astropart. Phys. 06 (2025) 007.

[177] P. A. Abell, J. Allison, S. F. Anderson, J. R. Andrew, J. R. P. Angel, L. Armus, D. Arnett, S. J. Asztalos, T. S. Axelrod *et al.* (LSST Science Collaborations), LSST Science Book, Version 2.0, arXiv:0912.0201.

[178] R. Laureijs, J. Amiaux, S. Arduini, J.-L. Augueres, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet *et al.*, Euclid Definition Study Report, arXiv:1110.3193.

[179] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith *et al.*, Array programming with NUMPY, Nature (London) **585**, 357 (2020).

[180] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright *et al.*, SCIPY 1.0: Fundamental algorithms for scientific computing in PYTHON, Nat. Methods **17**, 261 (2020).

[181] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin *et al.*, TENSORFLOW: Large-scale machine learning on heterogeneous systems (2015), 10.5555/3026877.3026899.

[182] A. Sergeev and M. D. Balso, HOROVOD: Fast and easy distributed deep learning in TENSORFLOW, arXiv:1802.05799.

[183] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, PYTORCH: An imperative style, high-performance deep learning library, arXiv:1912.01703.

[184] F. A. Torres, M. M. Negri, M. Inversi, J. Aellen, and V. Roth, Lagrangian flow networks for conservation laws, arXiv:2305.16846.

[185] A. Collette, *PYTHON and HDF5* (O'Reilly, 2013).

[186] J. D. Hunter, Matplotlib: A 2D graphics environment, Comput. Sci. Eng. **9**, 90 (2007).

[187] S. Fischbacher, T. Kacprzak, J. Blazek, and A. Refregier, Redshift requirements for cosmic shear with intrinsic alignment, J. Cosmol. Astropart. Phys. 01 (2023) 033.