Using Host Galaxy Photometric Redshifts to Improve Cosmological Constraints with Type Ia Supernovae in the LSST Era

Ayan Mitra1,2,3, Richard Kessler4,5, Surhud More1, and Renee Hlozek6

The LSST Dark Energy Science Collaboration

1 The Inter-University Centre for Astronomy and Astrophysics (IUCAA), Post Bag 4, Ganeshkhind, Pune 411007, India; ayanmitra375@gmail.com
2 Energetic Cosmos Laboratory, Nazarbayev University, Nur-Sultan 010000, Kazakhstan
3 Kazakhstan-British Technical University, Almaty, Kazakhstan
4 Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA
5 Department of Astronomy and Astrophysics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA
6 University of Toronto, 27 King’s College Circle, Toronto, ON M5S, Canada

Received 2022 October 14; revised 2022 December 5; accepted 2022 December 31; published 2023 February 27

Abstract

We perform a rigorous cosmology analysis on simulated Type Ia supernovae (SNe Ia) and evaluate the improvement from including photometric host galaxy redshifts compared to using only the “z_{spec}” subset with spectroscopic redshifts from the host or SN. We use the Deep Drilling Fields (~50 deg²) from the Photometric LSST Astronomical Time-Series Classification Challenge (PLAsTiCC) in combination with a low-z sample based on Data Challenge 2. The analysis includes light-curve fitting to standardize the SN brightness, a high-statistics simulation to obtain a bias-corrected Hubble diagram, a statistical+systematics covariance matrix including calibration and photo-z uncertainties, and cosmology fitting with a prior from the cosmic microwave background. Compared to using the z_{spec} subset, including events with SN+host photo-z results in (i) more precise distances for z > 0.5, (ii) a Hubble diagram that extends 0.3 further in redshift, and (iii) a 50% increase in the Dark Energy Task Force figure of merit (FoM) based on the w_0w_aCDM model. Analyzing 25 simulated data samples, the average bias on w_0 and w_a is consistent with zero. The host photo-z systematic of 0.01 reduces FoM by only 2% because (i) most z < 0.5 events are in the z_{spec} subset, (ii) the combined SN+host photo-z has ×2 smaller bias, and (iii) the anticorrelation between fitted redshift and color self-corrects distance errors. To prepare for analyzing real data, the next SN Ia cosmology analysis with photo-z should include non–SN Ia contamination and host galaxy misassociations.

1. Introduction

Since the discovery of cosmic acceleration (Riess et al. 1998; Perlmutter et al. 1999) using Type Ia supernovae (SNe Ia), this geometric probe has provided unique constraints on the dark energy equation of state (EOS) today, w_0, and its variation with cosmic time, w_a (Chevallier & Polarski 2001; Linder 2003). The most precise measurements of the dark energy EOS have been based on ~1000 spectroscopically confirmed SNe Ia with spectroscopic redshifts from the SN or host galaxy (Betoule et al. 2014; Scolnic et al. 2018; Abbott et al. 2019; Brout et al. 2022).

Over the next decade, much larger SN samples are expected from the Vera C. Rubin Observatory and Legacy Survey of Space and Time (LSST) and the Nancy Grace Roman Space Telescope. Spectroscopic resources will be capable of observing only a small fraction of the discovered SNe. To make full use of these future samples in cosmology analyses, well-developed methods have been used for photometric classification using broadband photometry (Lochner et al. 2016; Möller & de Boissière 2020). A photometric redshift method using the SN+host galaxy photo-z has been proposed (Kessler et al. 2010; Palanque-Delabrouille et al. 2010; Roberts et al. 2017), but a rigorous SN Ia cosmology analysis with photo-z has not been performed.

To analyze SN Ia samples with contamination from other SN types, the “BEAMS” framework was developed to rigorously use the photometric classification probabilities (Kunz et al. 2007; Hlozek et al. 2012). The BEAMS framework, combined with photometric classification, was first used to obtain SN Ia cosmology results from Pan-STARRS1 data (Jones et al. 2018). An extension to BEAMS, BEAMS with Bias Corrections (BBC; Kessler & Scolnic 2017, hereafter KS17), was used in Jones et al. (2018) and is currently used in the analysis of data from the Dark Energy Survey (DES; Vincenzi et al. 2023).

To analyze SN Ia samples using photometric redshifts, Kessler et al. (2010) and Palanque-Delabrouille et al. (2010) extended the SALT2 light-curve fitting framework (Guy et al. 2007) to include redshift as an additional fitted parameter and use the host galaxy photo-z as a prior. Dai et al. (2018) analyzed a simulated LSST sample including SNe Ia and SNe core collapse (CC) and applied both photometric classification (but not BEAMS) and the SALT2 photo-z method. They fit the resulting Hubble diagram with a flat ΛCDM model and recovered unbiased $\Omega_M$ with a statistical precision of 0.008. Using data from the DES, Chen et al. (2022) performed a...
photo-z analysis using a subset of $\sim$100 SNe Ia hosted by redMagic galaxies for which both photometric and spectroscopic redshifts are available. Fitting their Hubble diagram with a flat $w$CDM model, they find a $w$-difference of 0.005 between spectroscopic and photometric (redMagic) redshifts. Finally, Linder & Mitra (2019) and Mitra & Linder (2021) evaluated the impact of photometric redshifts for LSST using a Fisher matrix approximation that does not include light-curve fitting or bias corrections. They concluded that for $z < 0.2$, spectroscopic redshifts are necessary for robust cosmology measurements.

A hierarchical Bayesian methodology (2BEAMS; Roberts et al. 2017) has been proposed to combine photometric classification (BEAMS), photometric host galaxy redshifts, and incorrect host galaxy assignments. This method has been validated on a toy simulation of SN distances with random fluctuations, but the analysis does not include light-curve fitting, bias corrections, or systematic uncertainties.

Here we present a rigorous SN Ia cosmology analysis on simulated LSST data that includes host galaxy photo-zs and covers the first 3 yr. We use these photo-zs to include more distant SNe that would otherwise be excluded in a spectroscopically confirmed sample, and we evaluate the impact of including these additional SNe in the cosmology analysis. Our simulation is based on the cadence of the Deep Drilling Fields (DDFs) from the Photometric LSST Astronomical Time-series Classification Challenge (PLAsTiCC; Kessler et al. 2019a, see Section 3.1), combined with a low-$z$ sample based on the cadence of the Wide Fast Deep (WFD) fields. Our end-to-end analysis includes light-curve fitting, simulated bias corrections applied with BBC, a covariance matrix that includes systematic uncertainties, and fitting a bias-corrected Hubble diagram for cosmological parameters. We examine the $w$CDM and $w_0\Lambda$CDM models.

We adopt the photo-z method from Kessler et al. (2010), and we use the host galaxy photo-z as a prior. To focus on photo-z issues, we simulate SNe Ia only (without contamination) and assume that all host galaxies are correctly identified. Therefore, the BEAMS formalism is not used in the analysis. We use science codes from the publicly available SuperNova ANAlysis software package SNANA (Kessler et al. 2009), and we use the cosmology analysis workflow from Pippin (Hinton & Brout 2020).

This paper is presented as follows. In Section 2, we briefly review LSST and the Dark Energy Science Collaboration (DESC). Section 3 describes the simulated data sample, and Section 4 describes the cosmology analysis. Results are presented in Section 5, and we conclude in Section 6.

### 2. Overview of LSST and the DESC

The LSST is a ground-based stage IV DES program (Cahn 2009; Ivezić et al. 2019). It is expected to become operational in 2023 and will discover millions of SNe over the 10 yr survey duration. The Simonyi Survey optical telescope at the Rubin Observatory includes an 8.4 m mirror and a state-of-the-art 3200 megapixel camera (9.6 deg$^2$ field of view) that will provide the deepest and widest views of the universe with unprecedented quality. The LSST will observe nearly half the night sky every week to a depth of 24th magnitude in the six filter bands ($ugrizy$) spanning wavelengths from ultraviolet to near-infrared.

The DESC (Kessler et al. 2019a; Hložek et al. 2020) is under development to test early classification and to test processing large numbers of detection “alerts” expected from the Rubin Observatory.

### 3. Simulated Data

We do not work with simulated images; thus, we do not run the LSST difference imaging analysis (DIA) based on Alard & Lupton (1998). Instead, we simulate SN Ia light curves corresponding to the output of DIA and calibrated to the AB magnitude system (Fukugita et al. 1996). Following PLAsTiCC (Kessler et al. 2019a; Hložek et al. 2020), we use the cadence and observing properties from MINION1016, and we include a host galaxy photometric redshift and rms uncertainty based on Graham et al. (2018, hereafter G18), but we do not model correlations between the SNe and host galaxy properties. PLAsTiCC was designed to motivate the development of classification algorithms for photometric light curves from transients discovered by LSST.

PLAsTiCC included two LSST observing strategies: (1) five DDFs, covering $\sim$50 deg$^2$, that are revisited frequently and hence correspond to areas with enhanced depth, and (2) the WFD covering a majority of the southern sky (18,000 deg$^2$). We simulate a high-$z$ sample using DDFs and coadd the nightly observations within each band (Section 3.1). Since the PLAsTiCC DDF data have limited statistics at low redshifts, we compliment the PLAsTiCC data with a spectroscopically confirmed low-$z$ sample (Section 3.2) based on the WFD cadence used in DC2.

Rather than using the publicly available PLAsTiCC data, we regenerate the DDF simulation because our analysis needs a

### Table 1: Summary of Simulation Statistics

| $z$ Range | $N_{\text{gen}}$ Total$^a$ | $N_{\text{gen}}$ Trigger$^b$ | After $z_{\text{spec}}$ | Selection Cuts: $N_{\text{gen}}$ Full Sample |
|-----------|----------------------------|-------------------------------|--------------------------|----------------------------------------|
| Low-$z$   | 0.01–0.08                  | 4200                          | 696                      | 539                                    | 539                                    |
| High-$z$  | 0.03–1.55                  | 41,819                        | 12,808                   | 1482                                   | 4873                                   |

Notes.

$^a$ Total number of generated SNe Ia.

$^b$ Two or more detections separated by more than 30 minutes.

$^c$ Subset of events with spectroscopic redshift.

---

9 https://github.com/RickKessler/SNANA

10 6.7 m of effective aperture.

11 https://lsstdesc.org

12 https://github.com/LSSTDESC/dia_pipe

13 http://ls.st/Collection-4604
much larger sample for bias corrections that is not publicly available. We have verified that our new sample is statistically equivalent to the public data by comparing distributions of redshift, color, and stretch. Our simulation does not include contamination from core-collapse and peculiar SNe or DIA artifacts such as catastrophic flux outliers, point-spread function (PSF) model errors, and nonlinearities.

The simulation process adapted in this analysis is described in depth in Kessler et al. (2019b). To accurately measure biases on cosmological parameters, 25 statistically independent simulated data samples are generated, and each sample is analyzed separately.

A summary of the average simulation statistics is shown in Table 1. For the high-z sample, the number of generated events ($N_{\text{gen}}$ column of Table 1) is computed from the measured volumetric rate, the duration of the survey, and the 50 deg$^2$ area of the DDFs. For low-z, $N_{\text{gen}}$ is arbitrarily chosen such that the number of events after the selection requirements is roughly 500, which is about $\sim$10% of the high-z statistics. Examples of simulated light curves at different redshifts are shown by the black circles in Figure 1.

### Table 2

| Filter | WFD | DDF |
|--------|-----|-----|
|        | Depth $^a$ | Gap $^a$ | Depth | Gap |
| $u$    | 23.84 | 10.5 | 25.05 | 5.3 |
| $g$    | 24.80 | 11.9 | 25.52 | 7.3 |
| $r$    | 24.21 | 8.2  | 25.60 | 7.3 |
| $i$    | 23.57 | 8.6  | 25.19 | 7.3 |
| $z$    | 22.65 | 9.0  | 24.79 | 7.3 |
| $Y$    | 21.79 | 11.2 | 23.83 | 7.4 |

**Notes.**

$^a$ $5\sigma$ limiting magnitude.

$^b$ Average time (days) between visits, excluding seasonal gaps.

The original PLaSTic simulation covers the first 3 yr of LSST with 18 models that include both extragalactic and galactic transients. For this analysis, we simulate only SNe Ia using the SALT2 model (Guy et al. 2007). This model includes measured populations of stretch and color from Scolnic & Kessler (2016) with stretch– and color–luminosity parameters ($\alpha = 0.14$, $\beta = 3.1$), an intrinsic scatter of the model spectral energy distribution (SED), and a near-infrared extension (Pierel et al. 2018) to include the $i$, $z$, and $y$-band wavelength range. Correlations between SNe and host galaxy mass are not included. Next, the model SED is modified to account for cosmic expansion ($\Omega_M = 0.315$, $w = -1$, flatness), redshift, and Galactic extinction from Schlafly & Finkbeiner (2011). Filter passbands are used to compute broadband fluxes at epochs determined by the DDF cadence from OpSim.
To estimate the host galaxy photometric redshifts, PLAsTiCC used a color-matched nearest-neighbor (CMNN) photometric redshift estimator (G18). The CMNN uses a five-dimensional color space grid to train a set of galaxies and defines a distance metric that is used on the test set to assign the redshift and associated uncertainty. Figure 3(a) shows the photo-z residuals, $\Delta z_{\text{spec}} = z_{\text{phot}} - z_{\text{true}}$, as a function of $z_{\text{true}}$. To characterize the residuals, we follow Graham et al. (2018) and define metrics for an inner core resolution and outlier fraction using the quantity $\Delta z_{(1+z)} = |z_{\text{phot}} - z_{\text{true}}|/(1+z_{\text{phot}})$. The resolution is the width of the interquartile distribution of $\Delta z_{(1+z)}$ divided by 1.349 and denoted by $\sigma_{\text{IQR}}$. The outlier fraction ($f_{\text{out}}$) is the fraction of events satisfying

$$\Delta z_{(1+z)} > 3\sigma_{\text{IQR}} \text{ and } \Delta z_{(1+z)} > 0.06. \quad (1)$$

For $z_{\text{true}} < 0.4$, nearly all of the events have a spectroscopic redshift, and for $z_{\text{true}} > 1.4$, the SNe are too faint for detection. For the relevant redshift range ($0.4 < z_{\text{true}} < 1.4$), $\sigma_{\text{IQR}} = 0.025$ and $f_{\text{out}} = 0.13$.

Following PLAsTiCC, we use the volumetric rate model $R(z)$ based on Dilday et al. (2008) for $z < 1$ and Hounsell et al. (2018) for $z > 1$. The rate $R(z)$ we adopt is given by

$$R(z) = 2.5 \times 10^{-5}(1 + z)^{1.5} \text{yr}^{-1}\text{Mpc}^{-3} \quad (z < 1),$$

$$R(z) = 9.7 \times 10^{-5}(1 + z)^{-0.5} \text{yr}^{-1}\text{Mpc}^{-3} \quad (z > 1). \quad (3)$$

### 3.2. Low-z Data: Spectroscopic

We simulate a spectroscopically confirmed low-z sample based on the WFD cadence from DC2. We assume accurate spectroscopic redshifts and 100% efficiency up to redshift $z < 0.08$. The simulation code and SN Ia model are the same as for the high-z sample. Compared to DDFs, the WFD cadence has 30% fewer observations, on average, and a 1 mag shallower depth (Table 2).

### 4. Analysis

The SN Ia cosmology analysis steps are shown in Figure 4 and described below. This analysis is similar to the recent DC2 SN Ia cosmology analysis in Sánchez et al. (2022), except here we include DDFs and use photo-z information. The analysis is performed three times, each using the same low-z sample but varying the high-z data:

1. $z_{\text{phot}}$: full sample including both spectroscopic and photometric redshifts;
2. $z_{\text{spec}}$: subset with only accurate spectroscopic redshift from either the host galaxy or SN; and
3. $z_{\text{cheat}}$: full sample forcing $z_{\text{phot}} = z_{\text{true}}$.

#### 4.1. Light-curve Fitting

To standardize the SN Ia brightness, we fit each light curve to the wavelength-extended SALT2 light-curve model from Pierel et al. (2018), the same model used in the simulations. The SALT2 light-curve fit determines the time of peak brightness ($t_0$), amplitude ($a_0$), stretch ($x_0$), and color ($c$). Previous cosmology analyses have all used SNe with accurate $z_{\text{spec}}$; thus, redshift had always been a fixed parameter in the SALT2 fit. In our analysis, the SALT2 fit uses the methodology in Kessler et al. (2010) in which the redshift is floated as a fifth parameter, which we call “$z_{\text{phot}}$.”
Following Equation (1) in Kessler et al. (2010), the five-parameter SALT2 fit uses MINUIT (James & Roos 1975) to minimize the following $\chi^2$:

$$\chi^2 = \sum_i \left\{ \frac{(F_{i}^{\text{data}} - F_{i}^{\text{model}}(\bar{x}_5))^2}{\sigma_{F,i}^2} + 2 \ln(\sigma_{F,i}/\bar{F}_i) \right\} + \left\{ (\bar{z}_{\text{phot}} - \bar{z}_{\text{host}})^2 / \sigma_{\bar{z},\text{host}}^2 \right\},$$  

(4)

where $F_{i}^{\text{data}}$ is the SN flux for the $i$th observation, $F_{i}^{\text{model}}(\bar{x}_5)$ is the SALT2 model flux computed from the five SN-dependent parameters $\bar{x}_5 = \{t_0, x_0, x_1, \sigma_{\text{phot}}, \sigma_{\text{spec}}\}$, and $\sigma_{F,i}$ is the quadrature sum of statistical and SALT2 model uncertainties. The second term in Equation (4) accounts for the fact that the SALT2 model uncertainties depend on one of the fitted parameters ($\sigma_{\text{phot}}$) because of the dependence on rest-frame wavelength and epoch. A reference uncertainty ($\bar{F}_i$) is computed after the first fit iteration so that the $2 \ln(\sigma_{F,i}/\bar{F}_i)$ term is close to zero in the second and third fit iterations. The second row in Equation (4) is the host galaxy photo-$z$ prior, where $\bar{z}_{\text{host}}$ is the mean of the photo-$z$ probability density function (PDF), and $\sigma_{\bar{z},\text{host}}$ is the rms.

To estimate the initial SALT2 parameters prior to the fit, $\bar{z}_{\text{phot}} = \bar{z}_{\text{host}}$. The remaining initial parameter estimates are obtained by minimizing the $\chi^2$ in a very coarse grid search. After each MINUIT fit iteration, the wavelength range for each LSST passband is transformed to the rest frame using a fitted $\bar{z}_{\text{phot}}$. If the central rest-frame wavelength is outside of the SALT2 model range (2600–11000 Å), the passband is dropped in the next fit iteration; a previously dropped passband can be added if it is within the model range. If any passband is dropped or added, the fit iteration is repeated to ensure that a consistent set of passbands are included in the fit.

For the subset with accurate $\bar{z}_{\text{spec}}$, the redshift prior is so precise (0.0001) that such fits are essentially equivalent to fixing the redshift in a four-parameter fit. Note that $\bar{z}_{\text{phot}}$ refers to the fitted redshift for all events, including the $\bar{z}_{\text{spec}}$ subset.

### 4.2. Selection Requirements and Systematic Uncertainties

We apply the following selection requirements (cuts) based on analyses using real data:

1. at least three bands with maximum S/N $>4$,
2. successful light-curve fit,
3. $|x_1| < 3.0$,
4. $|\sigma| < 0.3$,
5. stretch uncertainty $\sigma_{x1} < 1.0$,
6. time of peak brightness uncertainty $\sigma_{t0} < 2.0$ days,
7. $P_{\text{fit}} > 0.05$,
8. $0.01 < \bar{z}_{\text{phot}} < 1.4$, and
9. valid bias correction (see Section 4.4).

The SALT2 light-curve fits for several events are shown by the smooth curves in Figure 1. After the selection requirements, the redshift distribution is shown in Figure 5(a) for the subset with and without $\bar{z}_{\text{spec}}$.

The $\bar{z}_{\text{phot}}$ residual versus $\bar{z}_{\text{true}}$ is shown in Figure 3(a) for all galaxies in the catalog and Figure 3(b) for host galaxies

---

**Figure 3.** Photo-$z$ residual ($\bar{z}_{\text{phot}} - \bar{z}_{\text{true}}$) vs. $\bar{z}_{\text{true}}$ for (a) the full host galaxy catalog, (b) a subset of the host galaxy catalog after trigger and cuts, (c) a SALT2-fitted SN-only photo-$z$ without a host galaxy prior, and (d) a combined SALT2-fitted SN+host photo-$z$ that is used for the Hubble diagram. Panels (b)–(d) have no $\bar{z}_{\text{spec}}$ events. The $\sigma_{\text{Phot}}$ and $f_{\text{out}}$ numbers in each panel are computed for $0.4 < \bar{z}_{\text{true}} < 1.4$. The source of the photo-$z$ is indicated on each vertical axis label.
after the SN Ia trigger and selection cuts. After the selection cuts, SNe associated with the host galaxy photo-
outliers tend to be excluded by the SALT2 fits and P \text{cut}; the core resolution is reduced by 10%, and the outlier fraction is reduced by 20%.

To compare the photo-\(z\) precision between the host and SN, we performed SALT2 light-curve fits without a host galaxy \(z\) prior to determine the SN-only \(z_{\text{phot}}\) residuals (Figure 3(c)); the SN-only \(z_{\text{phot}}\) core resolution is slightly (~1.1) better than for the galaxies in Figure 3(a), although the outlier fractions are the same. For the combined SN+host SALT2 fits, Figure 3(d) shows \(z_{\text{phot}}\) residuals versus \(z_{\text{true}}\); compared to fitting SNe only, the SN+host \(z_{\text{phot}}\) resolution is 30% smaller and has ~15% fewer outliers.

To evaluate systematic uncertainties, the SALT2 light-curve fits and BBC fit are repeated 17 times, each with a separate variation shown in Table 3. Each variation results in a distance modulus variation, and we compute a systematic covariance matrix \((\text{COV}_{\text{syst}})\) using Equation (6) in Conley et al. (2011).

We include variations in Galactic extinction, calibration, \(z_{\text{spec}}\), and host galaxy \(z_{\text{phot}}\). We do not include SALT2 modeling and training uncertainties, nor do we include uncertainties on the stretch and color populations.

Table 3

| Row | Label | Description | Value \(a\) |
|-----|-------|-------------|-------------|
| 1   | StatOnly | No systematic shifts | … |
| 2   | MWEBV | Shift \(E(B-V)\) | 5\% |
| 3   | CAL_HST | HST calibration offset | 0.007 \(\times \lambda\) |
| 4   | CAL_ZP | LSST zero-point shift | 5 mmag |
| 5   | CAL_WAVE | LSST filter shift | 5 Å |
| 6   | \(z_{\text{SPEC}}\) | Shift \(z_{\text{spec}}\) redshifts | \(5 \times 10^{-5}\) |
| 7   | \(z_{\text{PHOT}}\) | Shift \(z_{\text{phot}}\) redshifts | 0.01 |
| 8   | \(z_{\text{PHOTERR}}\) | Scale host \(z_{\text{phot}}\) uncertainty | 1.2 |

Note. 
\(^a\) Shift (or scale) applied to simulated data before each reanalysis.

For one of the 25 statistically independent samples, redshift-binned (solid black circles) and unbinned Hubble diagram from the BBC fit for \(z_{\text{spec}}\) (top) and \(z_{\text{phot}}\) (bottom) samples. Each bottom panel shows the Hubble residual \(\Delta \mu\) with respect to the true cosmology; the error bars show the rms in each BBC redshift bin (same redshift bins as in Figure 5).

For the seven systematic categories in Table 3, the ZP and WAVE rows each have one contribution per band, resulting in 12 band-dependent systematics. Adding the other five systematics rows results in a total of 17.
The galactic extinction uncertainty (row 2 in Table 3) is $\sigma_{\text{MWEBV}} = 0.05 \cdot \text{MWEBV}$ and taken from the Pantheon analysis (Scolnic et al. 2018). The Hubble Space Telescope (HST) calibration uncertainty (row 3) is from the DES SN Ia cosmology analysis (Table 4 in Brout et al. 2019) and based on Bohlin et al. (2014). The zero-point uncertainty (row 4) is from the LSST science road map (Section 3.3 in Ivezić & The LSST Science Collaboration 2018) and consistent with the Pan-STARRS 3π internal calibration accuracy (Schlafly et al. 2012; Magner et al. 2013). The wavelength calibration uncertainty (row 5) is from the Pantheon analysis in Scolnic et al. (2018).

The spectroscopic redshift uncertainty (row 6) is from Table 4 in Brout et al. (2019), which is based on low-redshift constraints on local density fluctuations (Calcino & Davis 2017). For the host galaxy photo-$z$ bias uncertainty (row 7), the statistical bias in our PLAsTiCC simulation is well below 0.01, as shown in the lower panel of Figure 2 in G18. This statistical bias is valid for the galaxy training set, but the bias for the subset of SN Ia host galaxies is likely to be larger. Without a photo-$z$ bias estimate for SN Ia host galaxies, we make an ad hoc estimate from the DES weak-lensing cosmology analysis in which Myles et al. (2021) found a statistical $z_{\text{phot}}$ bias of $\sim 0.001$, while their weighted $z_{\text{phot}}$ bias is 0.01, an order of magnitude larger. We use their weighted $z_{\text{phot}}$ bias of 0.01 as the systematic uncertainty. The uncertainty in the host $z_{\text{phot}}$ uncertainty (row 8) is from the variation in robust standard deviations in the upper panel of Figure 2 in G18.

### 4.3. Simulated Bias Corrections

To implement distance bias corrections in BBC (Section 4.4), we generate a large sample of $3.1 \times 10^6$ events (after cuts; Section 3.1), which consists of $2.6 \times 10^6$ high-$z$ events and $4.4 \times 10^5$ low-$z$ events. The bias correction is applied independently for high- and low-$z$, and thus the relative number of events in each subsample need not match the data. The simulation procedure is identical to that used for the simulated data, except for $\alpha$ and $\beta$. While fixed values are used for the data sample, a $2 \times 2$ $\alpha$, $\beta$ grid is used for the “biasCor” simulation to enable interpolation in BBC.

### 4.4. BEAMS with Bias Corrections

The BBC reads the SALT2-fitted parameters (high- and low-$z$) from the data and biasCor simulation and produces a bias-corrected Hubble diagram, both unbinned and in redshift bins. For each event, the measured distance modulus is based on Tripp (1998),

$$
\mu = m_B + \alpha x_1 - \beta c + M_0 + \Delta \mu_{\text{bias}},
$$

where $m_B = -2.5 \log_{10}(x_0)$, $\{\alpha, \beta, M_0\}$ are global nuisance parameters, and $\Delta \mu_{\text{bias}} = \mu - \mu_{\text{true}}$ is determined from the biasCor simulation in a five-dimensional space of $\{z, x_1, c, \alpha, \beta\}$. A valid bias correction is required for each event, resulting in a few percent loss. The distance uncertainty ($\sigma_\mu$) is computed from Equation (3) of KS17. Since there is no contamination from non--SNe Ia, all SN Ia classification probabilities are set to 1, and we do not use the BEAMS formalism.
There are two subtle issues concerning the use of $z_{\text{phot}}$ and its uncertainty $\sigma_z$. First, the calculated distance error from $\sigma_z$ ($\sigma_z^2$ in Equation (3) of KS17) is an overestimate because it does not account for the correlated color error that reduces the distance error. By floating $z_{\text{phot}}$ in the SALT2 fit, redshift correlations propagate to the other SALT2 parameter uncertainties; therefore, we set $\sigma_z^2 = 0$. The second issue concerns the $\mu_{\text{bias}}$ computation, where $\mu_{\text{true}}$ is computed at a SALT2-fitted $z_{\text{phot}}$ rather than the true redshift.

To avoid a dependence on cosmological parameters, the BBC fit is performed in 14 logarithmically spaced redshift bins. The fitted parameters include the global nuisance parameters ($\alpha, \beta, M_0$) and bias-corrected distances in 14 redshift bins. The unbinned Hubble diagram is obtained from Equation (5) using the fitted parameters.

If the same selection requirements are applied to each systematic variation for computing $\text{COV}_{\text{syst}}$, small fluctuations in the fitted SALT2 parameters and redshift result in significantly different samples, and these differences introduce statistical noise in $\text{COV}_{\text{syst}}$. We avoid this covariance noise by defining a baseline sample for events passing cuts without systematic variations, and we use this same baseline sample for all systematic variations. For example, if an event has a fitted SALT2 color parameter $c = 0.299$ and migrates to $c = 0.3001$ for a calibration systematic, this event is preserved without applying cuts that require $|c| < 0.3$.

To avoid sample differences from the valid bias-correction requirement, the BBC fit is run twice, and the second fit only includes events that have a valid bias correction in all systematic variations. Finally, for redshift systematics that result in migration to another redshift bin, the original (no syst) redshift bin is preserved for the BBC fit.

### 4.5. Cosmology Fitting and Figure of Merit

For cosmology fitting, we use a fast minimization program that approximates a cosmic microwave background (CMB) prior using the $R$-shift parameter (e.g., see Equation (69) in Komatsu et al. 2009) computed from the same cosmological parameters that were used to generate the SNeIa. The $R$ uncertainty is $\sigma_R = 0.006$, tuned to have the same constraining power as Planck Collaboration et al. (2021). We fit with $w\text{CDM}$ and $w_0w_a\text{CDM}$ models, where $w = [w_0 + w_a(1 - a)]$. The statistical+systematics covariance matrix is used. We fit both binned and unbinned Hubble diagrams.

For the $w_0w_a\text{CDM}$ model, the figure of merit (FoM) is computed based on the Dark Energy Task Force definition in Albrecht et al. (2006),

$$\text{FoM} \simeq \frac{1}{\sigma(w_0)\sigma(w_a)\sqrt{1 - \rho^2}},$$

where $\rho$ is the reduced covariance between $w_0$ and $w_a$.

### 5. Results

For one of the 25 statistically independent samples, we show the $z_{\text{spec}}$ and $z_{\text{phot}}$ Hubble diagram produced by the BBC fit, both binned and unbinned, in Figure 6. The Hubble residuals with respect to the true cosmology, $\Delta \mu = \mu - \mu_{\text{true}}$, are consistent with zero and do not show a redshift-dependent slope.

The BBC-fitted nuisance parameters are shown in Table 4 for the three analyses: $z_{\text{spec}}$, $z_{\text{phot}}$, and $z_{\text{cheat}}$. Averaging over the 25 samples, $\alpha$ and $\beta$ agree well with the simulated inputs. There is no true $\sigma_{\text{int}}$ for comparison, but we note that the $\sigma_{\text{int}}$ values agree well among the three analyses.

Next, we compare the BBC-fitted distance uncertainties ($\sigma_p$) in the redshift bins (Figure 5(b)). The $z_{\text{spec}}$ and $z_{\text{phot}}$ uncertainties are similar for $z < 0.5$, and at higher redshifts, the $z_{\text{phot}}$ uncertainty is significantly smaller than for $z_{\text{spec}}$. In addition to smaller distance uncertainties, the $z_{\text{phot}}$ redshift range extends $\sim 0.3$ beyond that of the $z_{\text{spec}}$ range.

At high redshift, the $z_{\text{cheat}}$ analysis shows little improvement over the $z_{\text{phot}}$ analysis. Defining an effective distance uncertainty per event in each redshift bin as $\sigma_p = \sigma_{\mu_{\text{spec}}} \times \sqrt{N}$, where $N$ is the number of events in the redshift bin, the $\sigma_p$ values for $z_{\text{cheat}}$ and $z_{\text{phot}}$ are the same to within a few percent. There are fewer $z_{\text{phot}}$ events (compared to $z_{\text{cheat}}$) because of selection cuts and unstable results between multiple light-curve fit iterations.

For the cosmology fitting, we fit the binned distances from the BBC fit and also performed unbinned fits to reduce the systematic uncertainty as described in Brout et al. (2021). While the unbinned cosmology fits result in smaller uncertainties, we find a significant bias that is driven by the calibration systematics. We have investigated numerical issues with covariance matrix inversion, speed-trick approximations in the cosmology fitting, and evaluation of derivatives $\text{COV}_{\text{syst}}$. We have not found an explanation of this bias; therefore, we present results only for binned distances.

For the subsections below, we define $w$ bias to be $w - w_{\text{true}}$, where $w$ is from the $w\text{CDM}$ cosmology fit. A similar definition is used for $w_0$ and $w_a$ for the $w_0w_a\text{CDM}$ model. The bias uncertainty is the standard deviation of the $w$-bias values divided by $\sqrt{25}$.

### 5.1. $w\text{CDM}$ Results

For the $w\text{CDM}$ cosmology fits, Table 5 shows the average $w$ bias and uncertainty among the 25 samples. The average $w$ bias is consistent with zero for both the $z_{\text{spec}}$ and $z_{\text{phot}}$ samples and with and without systematic uncertainties. The $w$-bias precision is $\sim 0.002$. The average $w$ uncertainty ($\langle \sigma_w \rangle$) for the $z_{\text{phot}}$ sample is 0.023 with systematics and is only slightly improved compared to $\langle \sigma_w \rangle = 0.025$ for the $z_{\text{spec}}$ sample. The additional sensitivity from the host galaxy $z_{\text{phot}}$ sample is small because the increased statistics are at higher redshifts, where the dark energy density fraction is much smaller compared to lower redshifts, where the sample is dominated by spectroscopic redshifts.

### 5.2. $w_0w_a\text{CDM}$ Results

For the $w_0w_a\text{CDM}$ model, the average bias, uncertainty, and FoM are shown in Table 6. While there was little improvement using the $z_{\text{phot}}$ sample with the $w\text{CDM}$ model, the $w_0w_a\text{CDM}$ improvement is much more significant because higher-redshift events, which are enhanced by the $z_{\text{phot}}$ sample, are more sensitive to evolving dark energy ($w_a$). With systematics, $\langle \text{FoM} \rangle = 95$ for the $z_{\text{spec}}$ sample and 145 for the $z_{\text{phot}}$ sample. The $w_{\text{spec}}-w_{\text{phot}}$ constraining power is shown in Figure 7 for a single simulated data sample.

The average bias is consistent with zero for both $w_0$ and $w_a$. For the $z_{\text{spec}}$ sample, the bias precision is $\sim 0.015$ and $\sim 0.07$ for $w_0$ and $w_a$, respectively. For the $z_{\text{phot}}$ sample, the bias precision is improved to $\sim 0.009$ and $\sim 0.04$. The $w_{\text{spec}}-w_{\text{phot}}$ average bias is
Notes.

a Average bias among 25 samples with uncertainty of std/√25.

b Average fitted uncertainty among 25 samples.

Table 6
Summary of w0\(w_a\)-CDM Cosmology Fits

| \(z_{\text{source}}\) | Syst. | \(\langle w_0 \text{ Bias}\rangle^a\) | \(\langle w_a \text{ Bias}\rangle\) | \(\langle \sigma_{w_0}\rangle^b\) | \(\langle \sigma_{w_a}\rangle\) | \(\text{FoM}\) |
|-----------------|-------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|----------------|
| \(z_{\text{spec}}\) | Stat. only | 0.0083 ± 0.0143 | −0.0658 ± 0.0674 | 0.076 | 0.353 | 136 |
| | Stat.+ syst. | 0.0067 ± 0.0140 | −0.0683 ± 0.0682 | 0.092 | 0.418 | 95 |
| \(z_{\text{phot}}\) | Stat. only | 0.0029 ± 0.0082 | −0.0228 ± 0.0342 | 0.048 | 0.211 | 237 |
| | Stat.+ syst. | 0.0011 ± 0.0091 | −0.0202 ± 0.0363 | 0.071 | 0.294 | 145 |

Figure 9. (a) \(\Delta \mu_\text{syst} - z\) vs. \(\Delta z_\text{syst} - z\) for a 0.01 systematic shift in host galaxy photo-z and linear fit with slope \(d\mu/dz\) in five \(z_{\text{spec}}\) bins (black circles with error bars) and the LCDM theory curve in red.

shown in Figure 8 and compared to the the \(w_0-w_a\) contours (statistical+systematic) for a single sample.

For the \(z_{\text{phot}}\) sample, the FoM averaged over 25 samples is \langle\text{FoM}\rangle = 237 with only statistical uncertainties and drops to \langle\text{FoM}\rangle = 145 when systematic uncertainties are included. Since there are many systematics contributing to the decrease in \langle\text{FoM}\rangle, we quantify the impact of each systematic \(i\) by recomputing the covariance matrix separately for each systematic \(\text{COV}_{\text{syst},i}\) and repeating the cosmology fit for each \(\text{COV}_{\text{syst},i}\). We finally compute the FoM ratios,

\[
\mathcal{R}_{\text{FoM},i} = \frac{\text{FoM}_{\text{syst},i}}{\text{FoM}_{\text{stat}}},
\]

where \(\text{FoM}_{\text{syst},i}\) is the FoM including only systematic \(i\), and \(\text{FoM}_{\text{stat}}\) is the FoM without systematic uncertainties. Note that \(\mathcal{R}_{\text{FoM},i} \leq 1\). Table 7 shows the \(\mathcal{R}_{\text{FoM},i}\), and the FoM degradation is dominated by the calibration systematics.

5.3. Discussion of Photo-z Systematics

The 0.01 photo-z shift systematic has a small (2%) effect on FoM for three reasons. First, the combined SN+host light-curve fit results in an average fitted redshift error of \(0.004\), or about half the host photo-z error. Second, this photo-z systematic does not affect \(z_{\text{spec}}\) events that dominate the lower-redshift region below about 0.5 (Figure 5(a)), and this \(z_{\text{spec}}\) region is most sensitive to redshift errors. The final reason is that the fitted \(z_{\text{phot}}\) and SALT2 color are anticorrelated; thus, a larger (smaller) \(z_{\text{phot}}\) results in bluer (redder) color, and this change in color self-corrects the distance error as illustrated in Figure 9.

To describe this distance self-correction, we first define \(\Delta z_{\text{syst}} - z\) as the difference between a SALT2-fitted \(z_{\text{phot}}\) with 0.01 host galaxy photo-z shift and nominal photo-z and similarly define \(\Delta \mu_{\text{syst}} - z\) as the distance difference from Equation (5). Figure 9(a) shows \(\Delta \mu_{\text{syst}} - z\) versus \(\Delta z_{\text{syst}} - z\) and a linear fit for the slope, \(d\mu/dz\), in one of five \(z_{\text{true}}\) bins. Figure 9(b) shows the measured \(d\mu/dz\) slope in five \(z_{\text{true}}\) bins (black circles) along with the LCDM theory curve in red. In the ideal limit, where the measured \(d\mu/dz\) exactly equals theory \(d\mu/dz\), the distance self-correction is perfect and results in no systematic uncertainty. Here the measured \(d\mu/dz\) are close to the theory curve; thus, the distance error is mostly corrected.

To gain further insight into the photo-z sensitivity, we first consider a naive systematic of shifting the fitted \(z_{\text{phot}}\) by 0.01 after the light-curve fit for the subset without a \(z_{\text{spec}}\). In this test, the compensating \(d\mu/dz\) points in Figure 9 are forced to be
zero, and there is no systematic reduction from a combined SN + host fit. Fitting the $w_0 w_a \Lambda$CDM model without COV$_{\text{syst}}$, the $w_0$ and $w_a$ biases are 0.03 and 0.15, respectively. Next, we consider the realistic case of shifting the host photo-$z$ before the SALT2 light-curve fit; the corresponding $w_0$ and $w_a$ biases are 0.001 and 0.003, more than an order of magnitude smaller than the naive systematic. While we have included an explicit host galaxy photo-$z$ systematic, there is no explicit analog for the SN. The SN photo-$z$ systematic is accounted for by the calibration and Galactic extinction contributions to the systematic uncertainty budget in Table 3, but it is difficult to untangle the impact of these systematics on distance and photo-$z$.

6. Conclusions

In this work, we presented cosmological dark energy constraints for simulated PLAsTiCC SN Ia data, and we continued the development of publicly available codes from SNANA and Pippin to analyze the data with a host galaxy photo-$z$ prior. For the $w_0 w_a \Lambda$CDM model, the dark energy FoM is $\sim237$ with only statistical uncertainties and drops to $\sim145$ with systematic uncertainties (Figure 7). This $z_{\text{phot}}$ FoM is 50% larger than the FoM obtained from the $z_{\text{spec}}$ subset that has a spectroscopic redshift from the host or SN. Averaging 25 independent data samples, the average bias on $w_0$ and $w_a$ is consistent with zero.

The systematic uncertainty from the host galaxy photo-$z$ results in only a 2% reduction in the FoM. This small impact is due to (i) nearly complete $z_{\text{spec}}$ at lower redshifts, (ii) smaller $z_{\text{phot}}$ bias from combining the SN and host, and (iii) anticorrelations between redshift and color that greatly reduce the distance error. While good $z_{\text{spec}}$ coverage is feasible for the DDFs, the WFD will likely have less $z_{\text{spec}}$ coverage and, using host galaxy photo-$z$s at lower redshifts, may increase the systematic uncertainty compared to this DDF analysis.

Simulated projections tend to be overly optimistic before a survey begins, particularly for the depth and average PSF. However, there are three key factors that are likely to improve future results: (1) here we simulated only 30% of the 10 yr baseline survey, (2) we used a CMB prior with constraining power to match Planck Collaboration et al. (2021) and did not assume improved CMB constraints during the LSST era, and (3) we did not include the ~50% FoM increase from fitting an unbinned Hubble diagram; this improvement awaits resolving the large $w_0 w_a$ bias associated with unbinned results.

Most SN Ia cosmology analyses over the past decade have used redshift-binned Hubble diagrams. These analyses include JLA (Betoule et al. 2014), Pantheon (Scolnic et al. 2018), PS1 single instrument (Jones et al. 2018), and DES (Abbott et al. 2019). The recent demonstration of smaller uncertainties with an unbinned Hubble diagram had not been rigorously tested until our analysis that shows biased cosmology parameters. We therefore encourage community effort to resolve this issue.

The next major effort is to develop the cosmology analysis for samples that include non–SN Ia contamination, host galaxy misassociation, and a more complete list of systematic uncertainties that includes host galaxy photo-$z$s models and intrinsic scatter of the SN brightness. Cosmology analyses using photometric classification and spectroscopic redshifts have been well developed on real data from PS1 (Jones et al. 2018) and DES (Vincenzi et al. 2023). Here we have developed and demonstrated a complimentary analysis using photometric redshifts and a spectroscopically confirmed sample.

Author contributions are listed below.

A. Mitra: co-lead project, SNANA simulations and analysis, writing
R. Kessler: co-lead project, software, analysis, writing
S. More: writing, review
R. Hlozek: development of PLAsTiCC.

A.M. acknowledges the funding of the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (grant No. AP08856149) and Nazarbayev University Faculty Development Competitive Research Grant Program No. 11022021FD2912. R.K. acknowledges pipeline scientist support from the LSST Dark Energy Science Collaboration. This work was completed in part with resources provided by the University of Chicago’s Research Computing Center.

This paper has passed an internal review by the DESC, and we thank the DESC internal reviewers: Dan Scolnic, Bruno Sanchez, and Martine Lokken.

The DESC acknowledges ongoing support from the Institut National de Physique Nucléaire et de Physique des Particules in France; the Science and Technology Facilities Council in the United Kingdom; and the Department of Energy, the National Science Foundation, and the LSST Corporation in the United States. DESC uses resources of the IN2P3 Computing Center (CC-IN2P3—Lyons/Villeurbanne—France) funded by the Centre National de la Recherche Scientifique; the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under contract No. DE-AC02-05CH11231; STFC DiRAC HPC Facilities, funded by UK BIS National E-infrastructure capital grants; and the UK particle physics grid, supported by the GridPP Collaboration. This work was performed in part under DOE contract DE-AC02-76SF00515.

ORCID iDs

Ayan Mitra @ https://orcid.org/0000-0002-9436-8871
Richard Kessler @ https://orcid.org/0000-0003-3221-0419
Surhud More @ https://orcid.org/0000-0002-2986-2371
Renee Hlozek @ https://orcid.org/0000-0002-0965-7864

References

Abbott, T. M. C., Allam, S., Andersen, P., et al. 2019, ApJL, 872, L30
Alard, C., & Lupton, R. H. 1998, ApJ, 503, 325
Albrecht, A., et al. 2006, arXiv:astro-ph/0609591
Betoule, M., Kessler, R., Guy, J., et al. 2014, A&A, 568, A22
Bohlin, R. C., Gordon, K. D., & Tremblay, P. E. 2014, PASP, 126, 711
Brout, D., Hinton, S. R., & Scolnic, D. 2021, ApJL, 912, L26
Brout, D., Scolnic, D., Kessler, R., et al. 2019, ApJ, 874, 150
Brout, D., Scolnic, D., Popovic, B., et al. 2022, ApJ, 938, 110
Cahn, R. N. 2009, From Quantum to Cosmos: Fundamental Physics Research in Space, ed. S. G. Turoshchev (Singapore: World Scientific), 685
Calcino, J., & Davis, T. 2017, JCAP, 2017, 038
Chen, R., Scolnic, D., Rozo, E., et al. 2022, ApJ, 938, 62
Chevallier, M., & Polarski, D. 2001, JMPD, 10, 213
Conley, A., Guy, J., Sullivan, M., et al. 2011, ApJL, 192, 1
Dai, M., Kuhlmann, S., Wang, Y., & Kovacs, E. 2018, MNRAS, 477, 4142
de Jong, R. S., Ageritz, O., Berbel, A. A., et al. 2019, Msngr, 175, 3
Delgado, F., Saha, A., Chandrasekharan, S., et al. 2014, Proc. SPIE, 9150, 915015

10
