Astro2020 Science White Paper

The Key Role of Supernova Spectrophotometry in the Next-Decade Dark Energy Science Program

Thematic Areas: ☐ Planetary Systems ☐ Star and Planet Formation

☐Formation and Evolution of Compact Objects ☒ Cosmology and Fundamental Physics

☐Stars and Stellar Evolution ☐Resolved Stellar Populations and their Environments

☐Galaxy Evolution ☐Multi-Messenger Astronomy and Astrophysics

Principal Author:
Name: Saul Perlmutter
Institution: UC Berkeley / LBNL
Email: saul@LBL.gov
Phone: (510) 486-5203

Co-authors: (names and institutions)
Greg Aldering (LBNL), Charlie Baltay (Yale), Susana Deustua (STScI),
Wendy Freedman (U. Chicago), Andy Fruchter (STScI), David Rubin (STScI),
Masao Sako (U. Penn), Nicholas Suntzeff (TAMU)

Abstract:

For LSST and WFIRST to together take a major step forward in the SN measurements of dark energy the leading systematics -- SN population drift and changes in dust properties -- both must be constrained. This requires space-based instrumentation to provide wide-wavelength-range spectrophotometry, not simply imaging, to measure these two effects.

Introduction

This white paper addresses the big step forward in the SN measurement of dark energy that WFIRST together with LSST can take, if high-quality spectrophotometric observations are available to keep systematic uncertainties as small as the projected statistical uncertainties.

For SNe, the primary limiting factor will, in fact, not be statistics: once LSST begins its cadenced observing, it will provide many tens of thousands of multi-band lightcurves for SNe Ia. With the repeated observations of the same fields these data sets are expected to have excellent calibration, and the observing depths (in the deep-drilling fields) are expected to be sufficient to ensure samples that are not suffering significant magnitude or color selection effects out to redshifts around $z = 0.8$. With such numbers, the limiting factors on the dark energy studies will clearly be (1) unconstrained systematic uncertainties, and (2) the inability to explore the higher redshifts and follow the SNe into the decelerating epochs both to improve the FoM and to look for time variation in dark energy properties. These are the two constraints that WFIRST's SN program can address, but this will require spectrophotometric follow-up capabilities.

In the big picture, the comprehensive next-generation study of dark energy that can begin to differentiate between dark energy theories is clearly such a challenging goal that it has long been recognized that this will require the combined use of the several known measurement approaches, including supernovae, baryon acoustic oscillations, and weak lensing, both to complement and to cross-check each other. When LSST and WFIRST are in operation, we expect to have next-generation data for all three of these measurement techniques, with DESI studying BAO from the ground, Euclid studying BAO and WL (and possibly SNe) from space, LSST studying WL and SNe from the ground, and WFIRST studying all three from space. These measurements will certainly be combined and used to check each other. So for any of these techniques to play a key role in the combined study of dark energy, that technique must individually make a big step forward with respect to preceding experiments using that same technique.

A principled approach to quantifying next-generation systematics

How do we specify scientific requirements for a next-generation SN measurement that will explore a level of accuracy where we do not yet know all the functional forms and values for the systematic uncertainties? It is not sufficient to simply design a measurement that must assume that all the properties of the supernovae and dust populations are simple gaussian or exponential distributions (and already known) and perhaps that they are constant in time -- and then simply *assume* these same population models in the analysis. In fact, we already have information on some of the complexity of these populations and their time dependence, so we can make some reasonable first estimates of the scale of the systematic changes that we need to design the measurement around.

For example, based on several hundred SN Ia spectrophotometric time series, we know that the SED behavior of the population cannot be described by just one or two principal components (as the currently standard SALT2 analysis does); The full range of actual SN Ia SED behavior in the data requires several more components to capture it (Saunders et al, 2018). There is already evidence that some of these components are more prominent in certain host environments, and other components in other environments. There is no reason to believe that the range of environments that we have studied at lower redshifts includes all of the properties of high redshift environments with sufficient representation that we would already have found all of the components of SN Ia SED behavior. We should therefore ask what additional uncertainty would be introduced to our analysis if one of the known SED components had died away in more recent SN Ia populations, so that we didn't have it as a known, a given, in our models. Rubin et al.'s UNITY analysis (see figure below) indicates that this becomes the dominant systematic uncertainty in the SN cosmology measurements, with the uncertainties due to some components matching or strongly exceeding that of the calibration uncertainties.

This consideration offers a reasonable way to quantify the need for a measurement plan that can handle evolutionary changes in the SN Ia SED population: if we have to allow the analyses to fit for a previously unknown SED component, how much does this reduce the cosmological power of different measurement plans? Would it still be a "next-generation" probe of dark energy? This drift was chosen to be within the known behaviors of SN Ia populations, so in that sense it is a modest proposed systematic evolutionary change, and when we explore the unknown redshifts we could well be seeing new changes outside of such current known behaviors. It seems reasonable -- and important -- to assume at least this degree of SED population drift.

Similarly, we already have information on the range of dust properties in different galaxies and within galaxies. It is known that the ratio of selective extinction, R_V, varies from 1.3 to 3.1 in different environments (e.g., Patat et al., 2015, Schlafly et al., 2016, Huang et al., 2017, Cikota et al., 2018), and that some of these are preferentially younger environments (Sullivan et al., 2010). There is also evidence, particularly from more highly reddened SNe (where individual SNe yield strong R_V constraints), that this range of R_V is being probed by SNe Ia. (As discussed below, Rubin et al.'s UNITY analysis indicates that the systematic uncertainty associated with this R_V drift is comparable to that of the calibration uncertainties.) Again, it seems reasonable to assume that this degree of R_V drift must be fit in the analysis, and we need to know how this reduces the cosmological power of different measurement plans.

With a similar approach to setting the scale for modest "known unknown" systematics (and setting aside for the moment the more dramatic possible systematics outside of these known ranges of behavior), we can give a partial list of such effects that should be included in a cosmology analysis. These include fundamental photometry calibration, flat-field/dark sub, scattered light, K-corrections, S-corrections, selection biases, count-rate nonlinearity, change of mean dust E-B with z, change in R_V with z, Milky Way extinction zero point, normalization, and R_V, intergalactic extinction, changes in SN Ia colors with z, host correlation changes with z, M_*, SFR, and hidden parameter selection bias. So far, many of these have been included in the comprehensive LSST+WFIRST simulation/analysis codes, while others have been studied separately and in various combinations. Each systematic uncertainty that is considered reduces the cosmological power of the measurement, and lowers the FoM, and none of the ones considered has any reason to cancel the effect of the others, so together the FoM would only be lower when they are combined.

Results with this approach to systematics, with and without spectrophotometric data

David Rubin's "UNITY" analysis (Rubin et al, 2015) has shown, with a hierarchical Bayesian approach, how to include the wide variety of systematic uncertainties together with multiple data sets in a single fit. This analysis then allows direct comparisons of the importance of the systematics to the cosmology fits, and several sources of systematics are clearly prominent at the level of precision that LSST and WFIRST are pursuing: Although up until recently calibration has a been considered to be the largest contributor to the uncertainty (Kessler et al. 2009; Conley et al. 2011; Suzuki et al. 2012; Betoule et al. 2014; Scolnic et al. 2014, Scolnic et al. 2018), with these new, more ambitious data sets (targeting improved calibration) the contribution from the evolutionary drift of R_V would be as large or larger, and a drift among the known population components of the SN Ia's SED is likely to dominate over both calibration and R_V if not controlled. (The following figure from Rubin et al shows the relative contributions of these systematics contributions.)

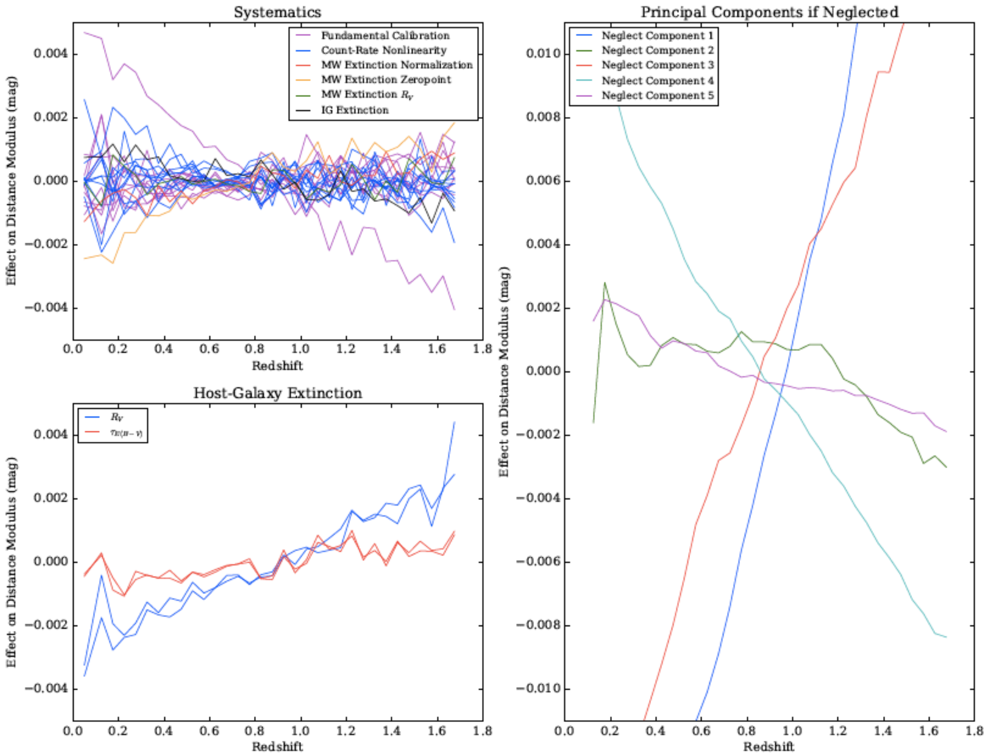

Figure 7. Contribution to the distance modulus uncertainty as a function of redshift for key parameters in the analysis (shown here for the Space-High survey). We define this as $(\partial\mu/\partial p)\sigma_p$ for each parameter p, then assume Gaussian uncertainties, and approximate as $C_{\mu,p}/\sigma_p$ (where $C_{\mu,p}$ is the covariance between each distance modulus bin and the parameter). For clarity, we subtract the mean in redshift from each term, as the additive scale is degenerate with the absolute magnitude, and is thus not important for our analysis. Thus, the count-rate nonlinearity impact is not zero for the nearby ground-observed SNe. Top left panel: parameters representing systematics in the analysis. Bottom left panel: parameters describing the dust extinction distribution as a function of redshift. Right panel: effect of neglecting one of the principal components when computing distance moduli, assuming that the mean changes by 1σ . The scales are the same in all panels, showing that many systematics enter the distance moduli at the same scale.

Spectrophotometric observations of the SNe makes it possible to constrain the latter two dominant systematic uncertainties -- and to do so even at redshifts between z = 1 and 2. By directly measuring wide-wavelength-range spectrophotometry, it is easy to determine differences in the spectral components making up the population of supernovae at different redshifts, something that is not possible with photometric observations. (We note that the resolution and signal-to-noise of the proposed WFIRST spectrophotometric SN programs can be used to match like-to-like, "twin" high- and low-redshift SN spectrophotometry. The current baselined WFIRST prism is aimed at accomplishing this for SNe up to redshifts of z ~ 1.2, while the previously studied WFIRST IFC spectrograph could accomplish this up to z ~ 1.8.) With such spectrophotometry it is also possible to separate out spectroscopically the intrinsic color of each SN (e.g., constraining intrinsic B-V by measuring silicon velocity), making it possible to also fit R_V as a function of redshift.

Without the spectrophotometry to constrain these two major sources of systematics that need to be considered, the cosmology fits would have to marginalize over a prior with a reasonable range of known R_V differences and known SN SED differences. (Even this would not be a conservative error bar, however, since we would not know that the actual drift is within this range.) The result of such a (prior-dependent) fit is a much weaker constraint on cosmology than if actual spectrophotometric measurements of the drifting R_V and SED are performed, so the FoM drops greatly. For example, studies of WFIRST planned programs show that fitting an unknown SED component *using* good spectrophotometry together with wide-field imaging affects the FoM very little, but *without* good spectrophotometry, the FoM drops dramatically.[1]

For a simple example of this, we focus on just one feature among all the SED features that are used in twinning and to constrain SED evolutionary drift, the Si II velocity, since it has been extensively discussed in the literature and provides an easier test case. Simulations of observations obtainable with an IFC on WFIRST with the program's planned exposure times provides a measurement of the Si velocity with a dispersion of approximately 700 km/sec, good enough to differentiate the known difference among SN populations. Dropping this measurement of the Si velocity would result in a dramatic reduction in the figure of merit, from, e.g., FoM = 362 to 98, a reduction of over 70%. (The current WFIRST prism will allow a measurement between these extremes, since it provides spectrophotometry for more than half of the redshift range; these estimates will soon be possible as the prism specifications are now being established.)

Previous studies of FoM have either neglected these key systematics or assumed very simple models of the dust and supernova behavior and then fit for those assumed models. Such studies then falsely appear to give a comparable FoM for photometry-only programs and spectrophotometric programs. (Since they are so constrained by model assumptions they would also predict an excellent FoM for an LSST-alone program.)

Only two techniques have so far been shown to improve the SN Ia dispersion significantly, and hence narrow down the range of systematic uncertainties: (1) spectrophotometric matching of "twin" supernovae, and (2) rest-frame near-IR photometry, in the rest-frame H band particularly. The latter technique cannot be extended to redshifts significantly above 0.3, even with WFIRST photometry, because the reddest photometry available will then be bluer than the rest-frame H band; beyond z = 0.8 the reddest photometry is in rest-frame optical bands. While these bands do help constrain dust (and rest-frame Y band is particularly helpful, as it is available below z = 0.8), they do not achieve the major step forward in tight constraints on systematics that the "twinning" or H band do. Not having a way to accomplish this systematics goal, neither LSST nor WFIRST without spectrophotometric capabilities is competitive with WFIRST with such

---

[1] In the scenarios that were studied, fitting an unknown SED component *using* IFC spectrophotometry together with wide-field imaging, reduces the DETF figure of merit (FoM) from 362 to between 346 and 361 (that is, dropping by between <1% and 4%), depending on which known component is assumed missing; but *without* any spectrophotometric observations, the FoM drops to between 28 and 252 (that is, loss of between 30% and 90% of its value).

capabilities: the FoM without the $z > 0.9$ program is a significant reduction, and the SN study of time variation of dark energy properties at higher redshift is lost completely.

The two primary sources of systematic uncertainty, SED and dust drift, also cannot be constrained by simply substituting ground-based spectroscopy for space-based spectrophotometry. The S/N and quality (after host-galaxy subtraction) of these ground-based spectra have never yet succeeded in making strong low-to-high-redshift comparisons, even though many papers have tried to do less ambitious matching than the "twins" technique. (Aside from the unsolved problems of spectrophotometry and host subtraction in ground-based data, at redshifts above $z \sim 1$ the available S/N from the ground is, moreover, below the level that has been shown to be needed for "twinning"; It is thus below the level needed to systematically detect drifting SED populations or to twin high-redshift SNe Ia to low-redshift matches.) In addition, ground-based spectroscopy is extremely inefficient: even 100 dark nights of Subaru's multi-fiber PFS instrument only nets ~350 SN spectra that are just good enough for typing and redshift determination.

Finally, it is important to note that these next-generation measurements of dark energy are exploring a new discovery space, both in precision and in redshift. Our experience has been that every new exploration finds new, previously unconsidered effects that have to be studied and controlled for, if a precision measurement is to be successful. Otherwise, at best one is aware of disagreements between measurement techniques, but then there are no tools to understand why. This white paper has focussed on the *known* systematics, but exploring into the unknown with no tools to address unknown systematics would be foolhardy. To put these risks in more day-to-day terms: Launching a WFIRST mission without the capability to measure the known range of SN Ia SEDs and dust R_V values is like going on a trip to the outback without bringing antibiotics for a known, potentially deadly infection you already have; launching WFIRST without the capability to explore SED and dust behavior outside this range is like canceling your health insurance to save money because, aside from this infection, you are basically healthy right now.

Conclusion

The bottom line is that for the next big observing programs, LSST and WFIRST, to together take a major step forward in the SN measurements of dark energy the two leading systematics -- due to evolutionary drift within known dust properties and within the known range of SN Ia SEDs -- both must be constrained. And these constraints must be useable for the full range of redshifts, so that the SN measurement can be taken well beyond the redshift range of the ground-based observations. This requires the space-based instrumentation to provide the wide-wavelength-range spectrophotometry to measure these two effects.

The whole philosophy of the upcoming generation of dark energy measurements calls for the kind of care only made possible by using photometry together with spectrophotometry measurements of SNe. *There is no point in a program of measurements that would only be believed if it finds an expected result like $w = -1 =$ constant. We must be able to trust a surprising result, or else we are just setting ourselves up for confirmation bias.*

References

Betoule, M., Kessler, R., Guy, J., et al. 2014, A&A, 568, A22
Cikota, A., Hoang, T., Taubenberger, S., et al. 2018, A&A, 615, A42
Conley, A., Guy, J., Sullivan, M., et al. 2011, ApJS, 192, 1
Huang, X., Raha, Z., Aldering, G., et al. 2017, ApJ, 836, 157
Kessler, R., Becker, A. C., Cinabro, D., et al. 2009, ApJS, 185, 32
Patat, F., Taubenberger, S., Cox, N. L. J., et al. 2015, A&A, 577, A53
Rubin, D., Aldering, G., Barbary, K., et al. 2015b, ApJ, 813, 137
Rubin, D., et al. in prep.
Saunders et al., The Astrophysical Journal, Volume 869, Issue 2, article id. 167, 25 pp. (2018)
Schlafly, E. F., Meisner, A. M., Stutz, A. M., et al. 2016, ApJ, 821, 78
Scolnic, D., Rest, A., Riess, A., et al. 2014, ApJ, 795, 45
Sullivan, M., Conley, A., Howell, D. A., et al. 2010, MNRAS, 406, 782
Suzuki, N., Rubin, D., Lidman, C., et al. 2012, ApJ, 746, 85