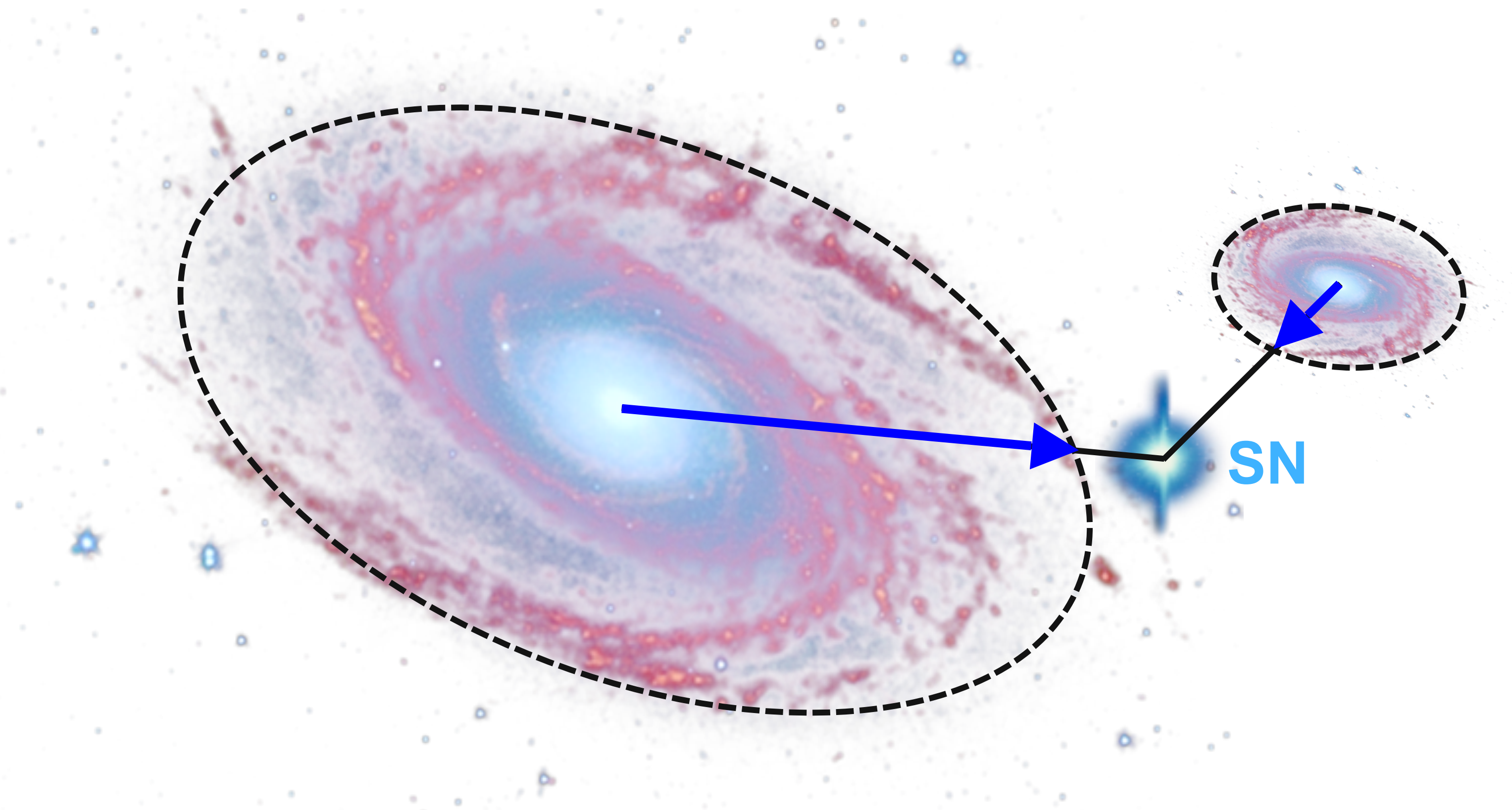
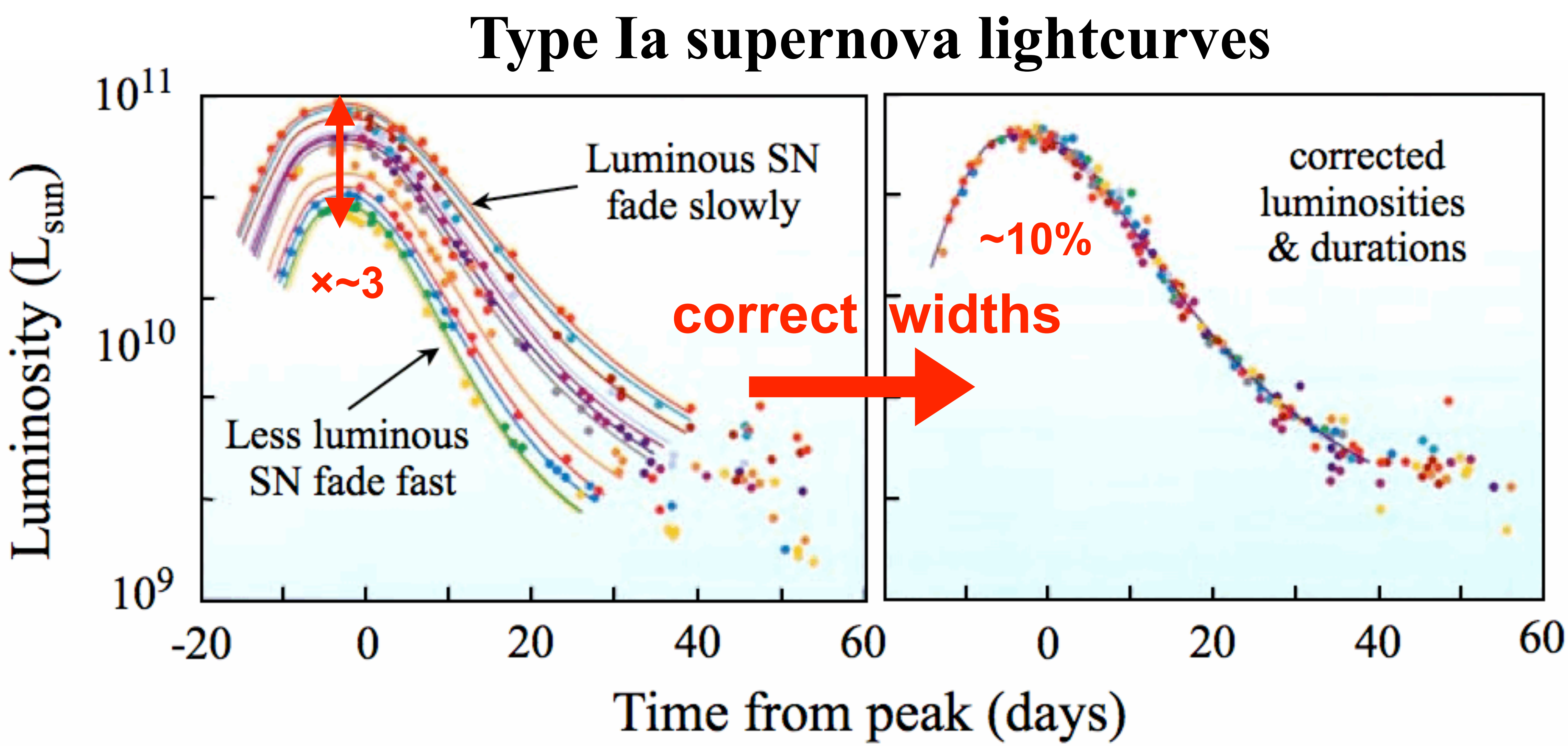
Host Galaxy Identification for Supernova Surveys



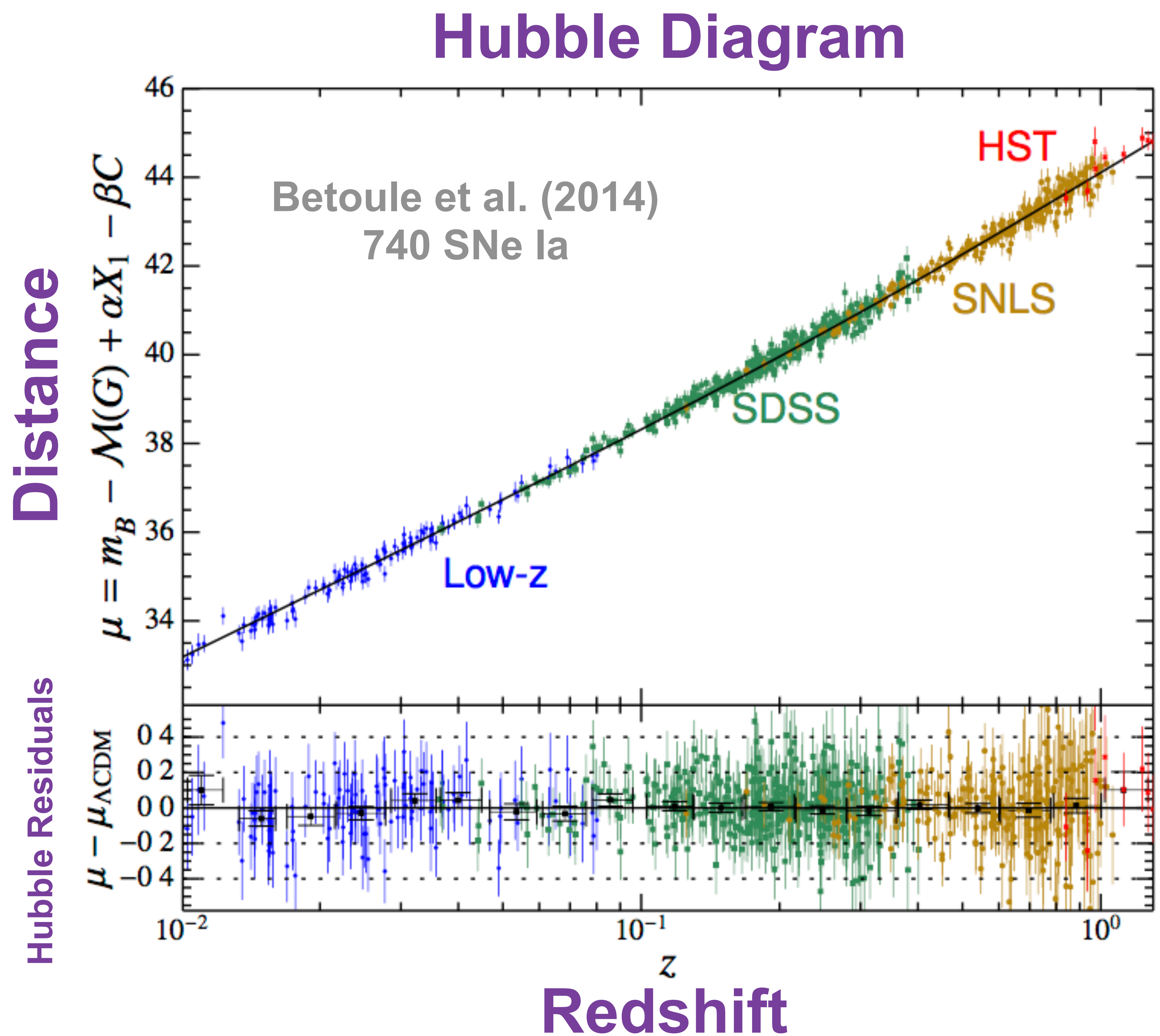
Ravi Gupta (Bldg 360, Room L-173)
Steve Kuhlmann, Eve Kovacs, Hal Spinka, Camille Liotine, Kasia Pomian + DES
HEP Division Young Scientist Symposium Series
16 February 2016

Type Ia Supernovae (SNe Ia): Standardizable Candles

- SNe Ia are thermonuclear explosions of carbon-oxygen white dwarfs
- Peak luminosity is related to both lightcurve width and color
- Calibrating the luminosity based on these empirical relations allows us to use SNe Ia as distance indicators and probe cosmology via the distance-redshift relation



Type Ia Supernovae (SNe Ia): Cosmological Probes

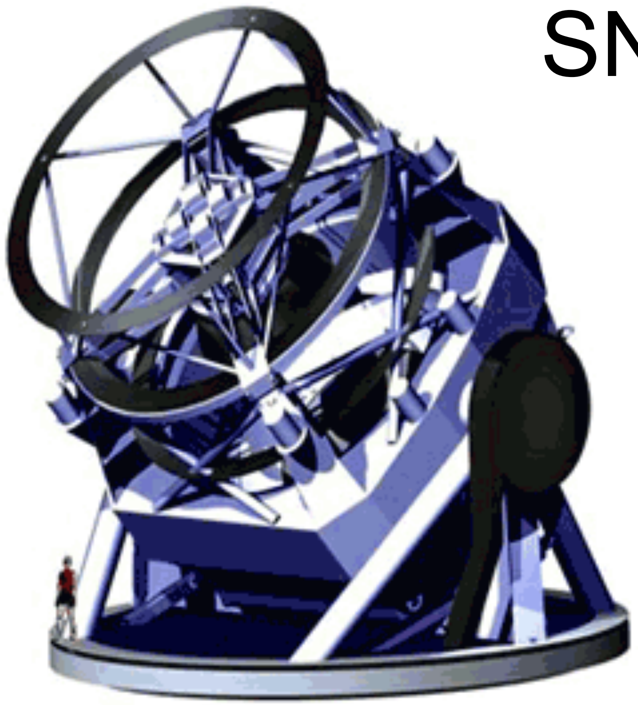


- Past surveys discovered tens to hundreds of SNe Ia
- Current and future surveys will find thousands and even hundreds of thousands more
- SN cosmology is becoming limited by systematic uncertainties rather than statistics:
  1. Photometric calibration
  2. Host galaxy correlations

The Importance of Host Galaxy Identification

- *Host galaxy identification* ("host matching") is a crucial step for modern SN surveys
- The Dark Energy Survey (DES) is on track to discover ~3500 SNe Ia
- The upcoming Large Synoptic Survey Telescope (LSST) will discover ≫10K SNe Ia
- In the absence of SN spectroscopy to determine SN types, we **rely mainly on host galaxy spectra** to obtain redshifts which are used to photometrically type SNe

LSST: ≫10K SNe Ia

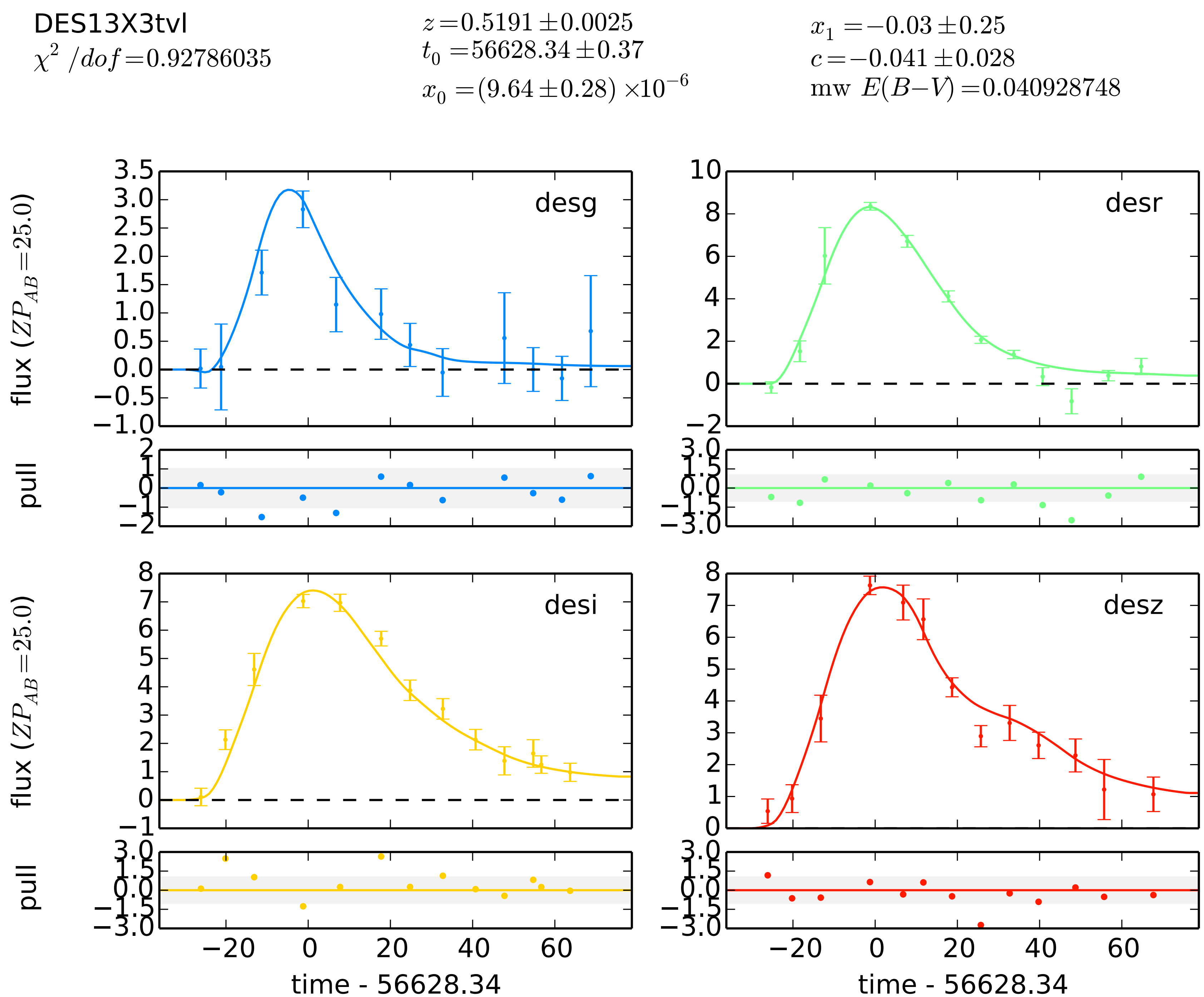


LSST Corp.

The Importance of Host Galaxy Identification: Photometric SN Classification

- By fitting the shape of the lightcurve, we can determine if the SN was Type Ia or other type
- Redshift of host galaxy (from spectrum of the host) greatly improves fit
- Only ~10% of our final sample of SNe Ia will be spectroscopically confirmed
- The majority rely on this method of photometric classification

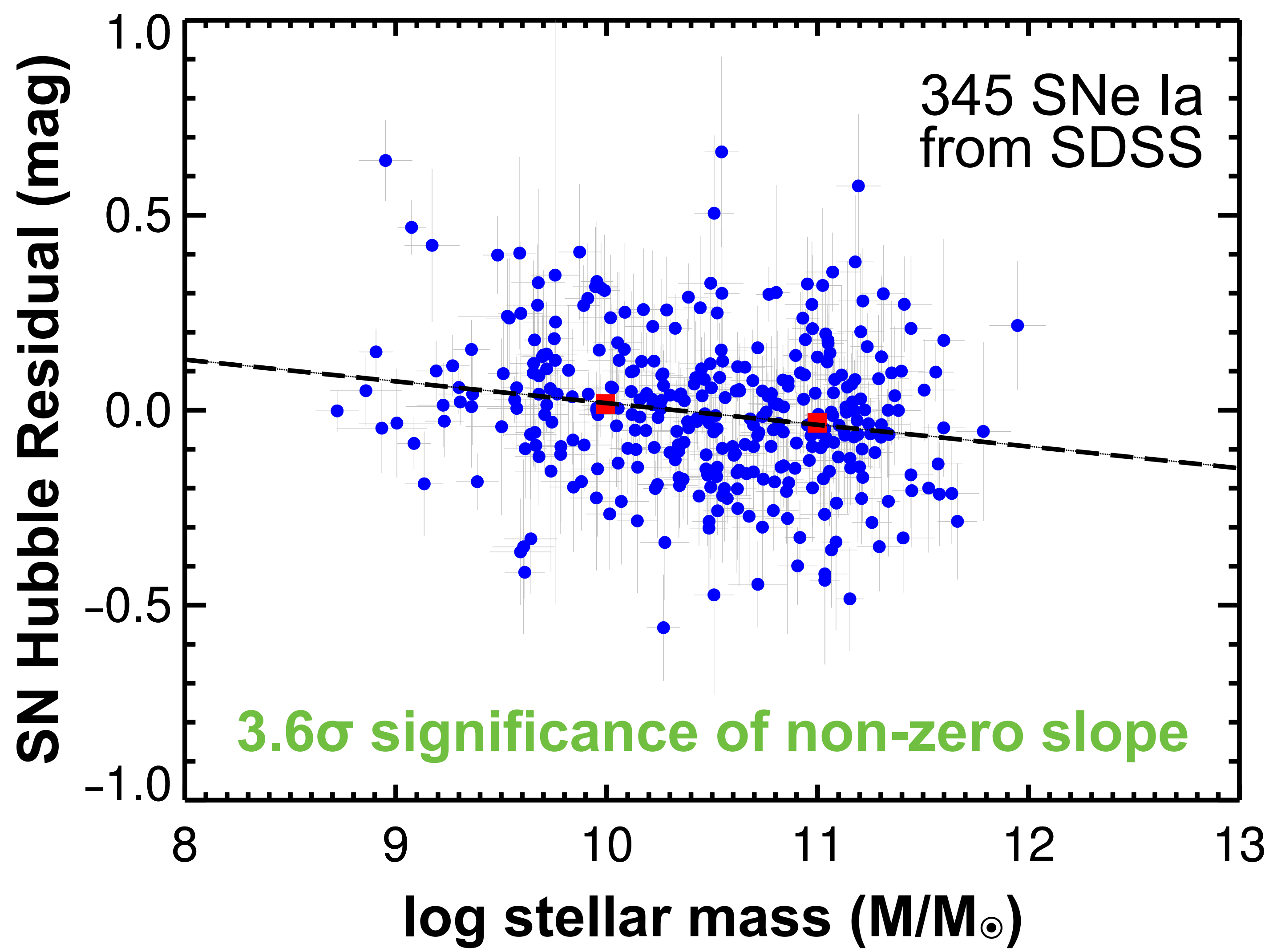
DES Year 1 SN Ia candidate fit using host galaxy redshift prior



The Importance of Host Galaxy Identification: Host Galaxy Correlations

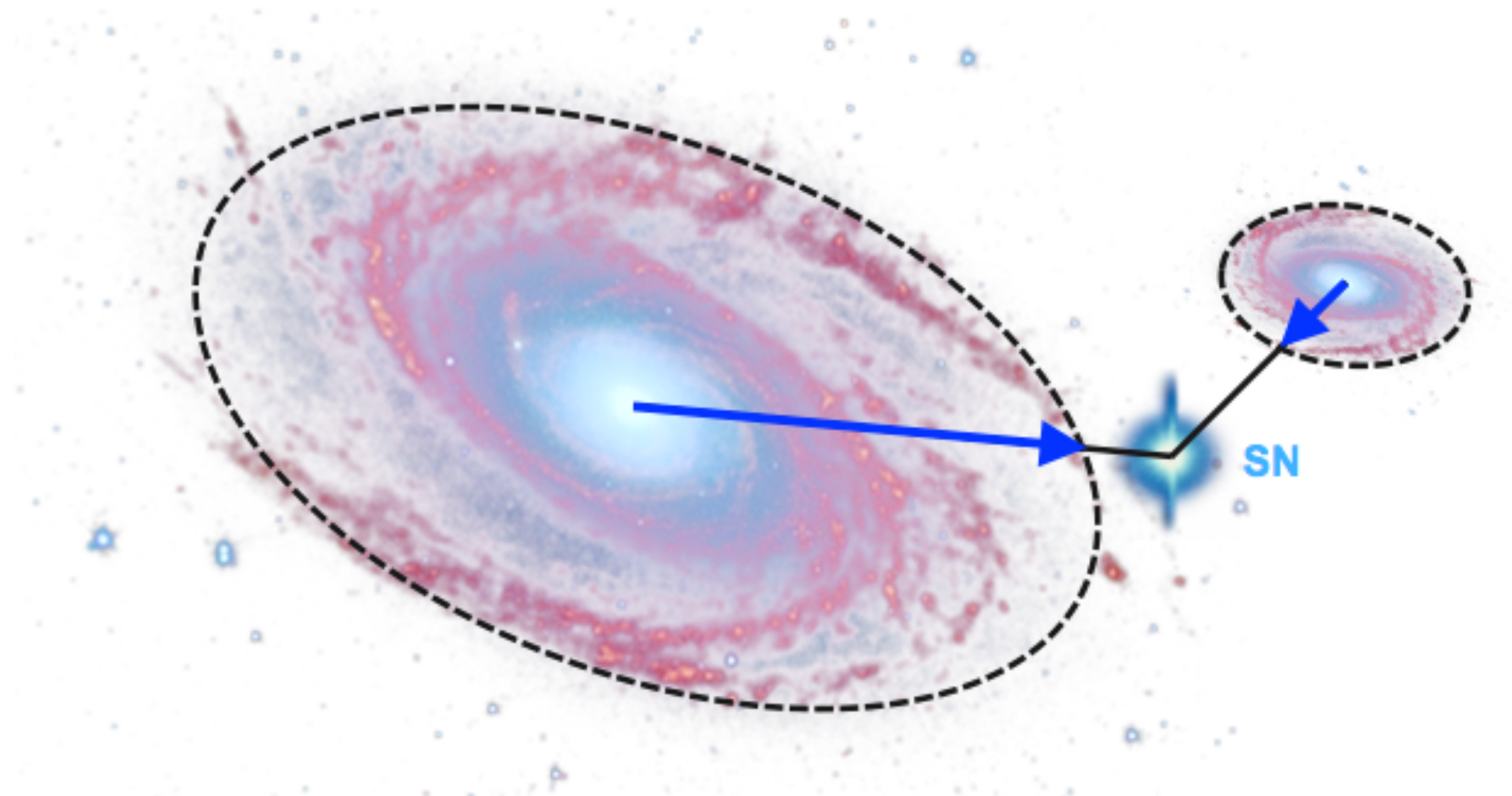
- In addition, SN luminosities are known to correlate with host galaxy properties
- The origin of this correlation is not yet understood, but cosmology analyses already use host galaxy properties to correct for SN luminosities
- **Reliable identification of host galaxies is essential for cosmology and SN science**

Wolf et al. (2016, ApJ accepted)



Method: Directional Light Radius (DLR)

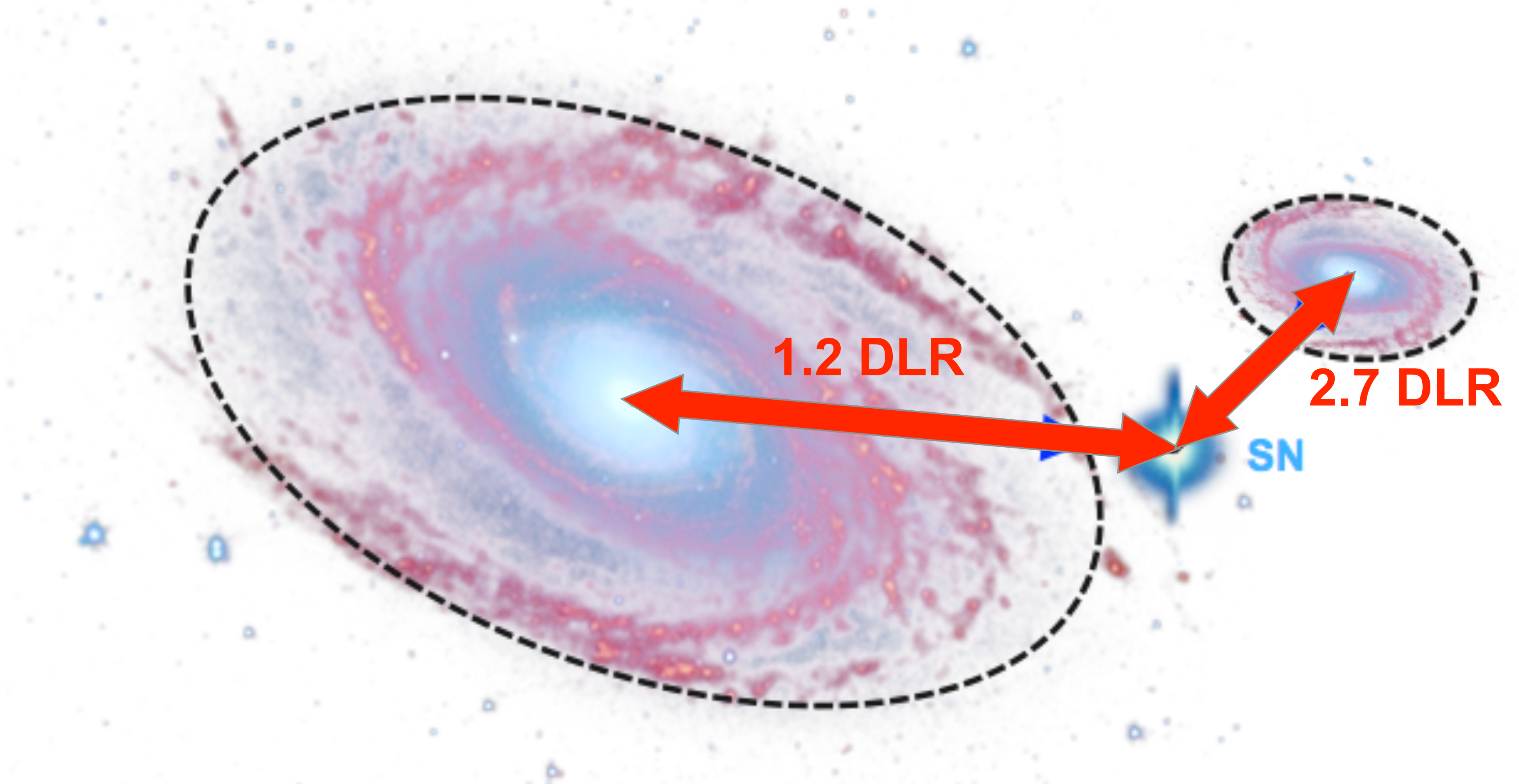
DLR = radius of a galaxy in the direction of the SN



Method: Directional Light Radius (DLR)

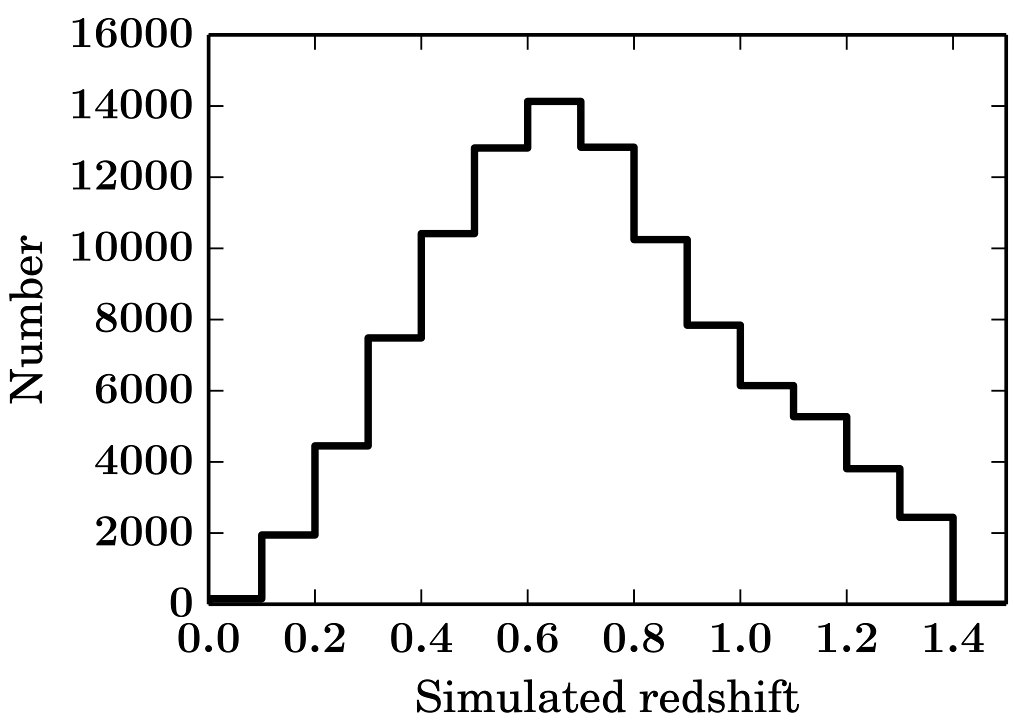
DLR = radius of a galaxy in the direction of the SN

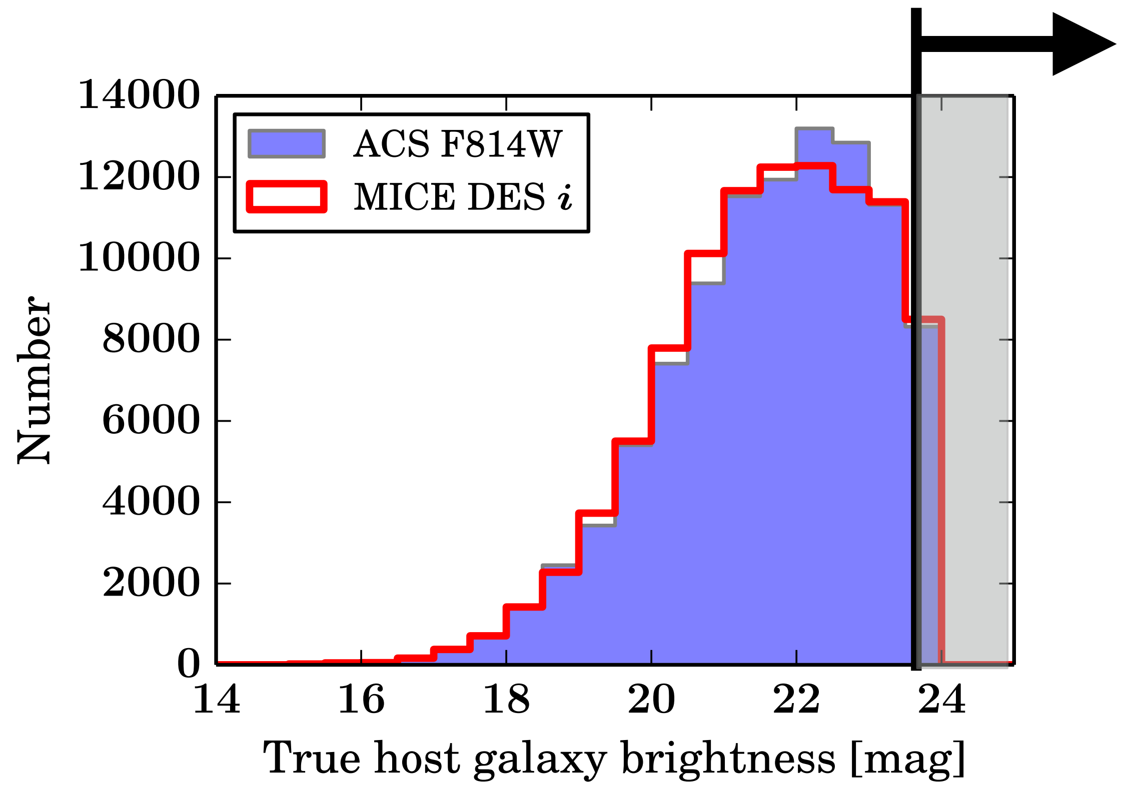
- Search for galaxies within a 30″ radius of SN position
- Match SN to the galaxy that is nearest in units of galaxy radius (DLR)
- In this way, the separation distance is normalized by the apparent size of the galaxy



Testing Host Matching with Simulations

- Using galaxy catalogs we can place simulated SN locations onto galaxies
  - Mock catalog: MICECATv2.0 (Carretero et al. 2015)
  - Real catalog: *HST* ACS General Catalog (Griffith et al. 2012)
- SN redshifts simulated to reproduce DES-like SN Ia redshift distribution
- SN positions placed such that they follow the light distribution of their host galaxies (Sérsic profiles)
- When matching, assume fiducial hostless SN rate of 5% to simulate magnitude-limited survey

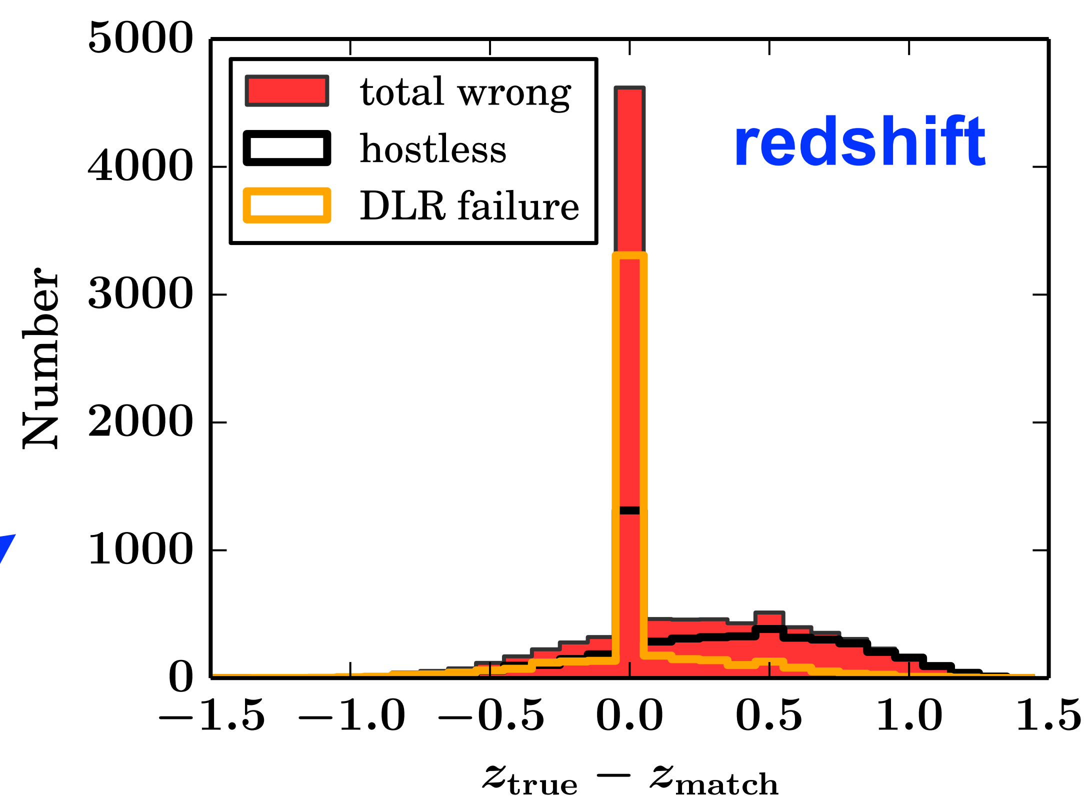


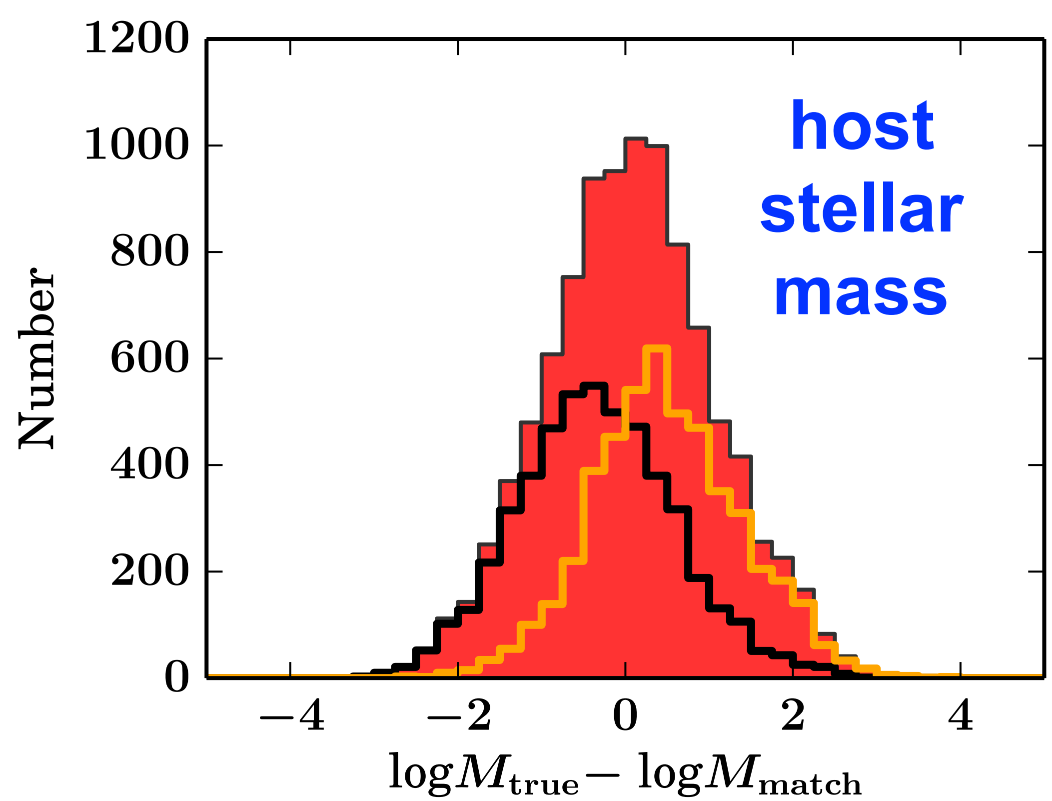


HOSTLESS: After simulating SNe, perform a magnitude cut on catalog such that the faintest 5% of hosts are removed

Testing Host Matching with Simulations: Results

- DLR distance to nearby galaxies is computed from the galaxy coordinates, sizes, shapes, and orientations given in the catalog
- Nearest galaxy in DLR-space is designated as the host
- Since the true host is known, we can test the matching accuracy
- **The DLR method performs with ~91% accuracy** (we know the 5% hostless will be mismatched to galaxies brighter than the true host)



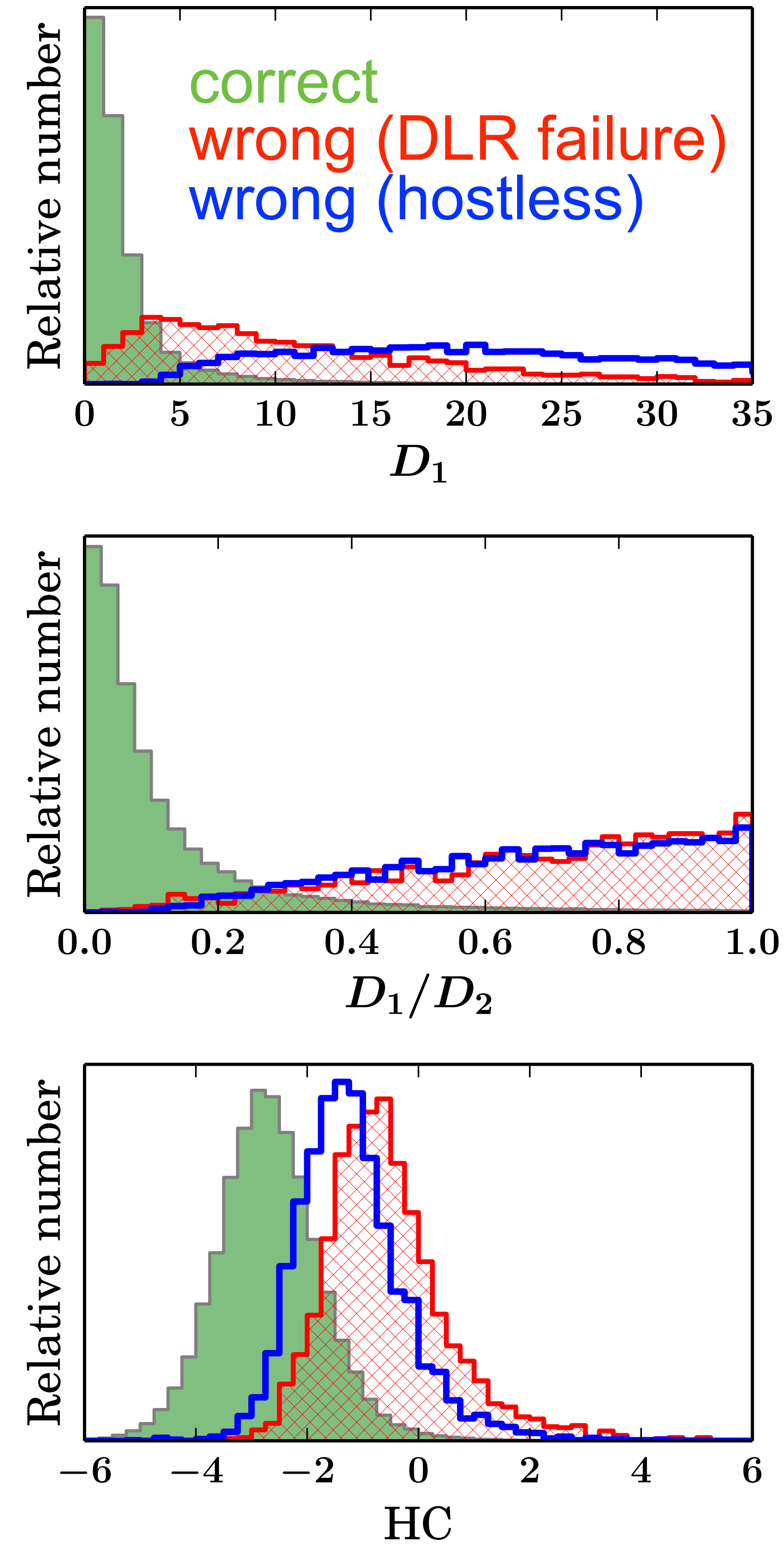


Even if the incorrectly-matched host happens to have a similar redshift,…

… this is still a problem given what we know about SN-host correlations

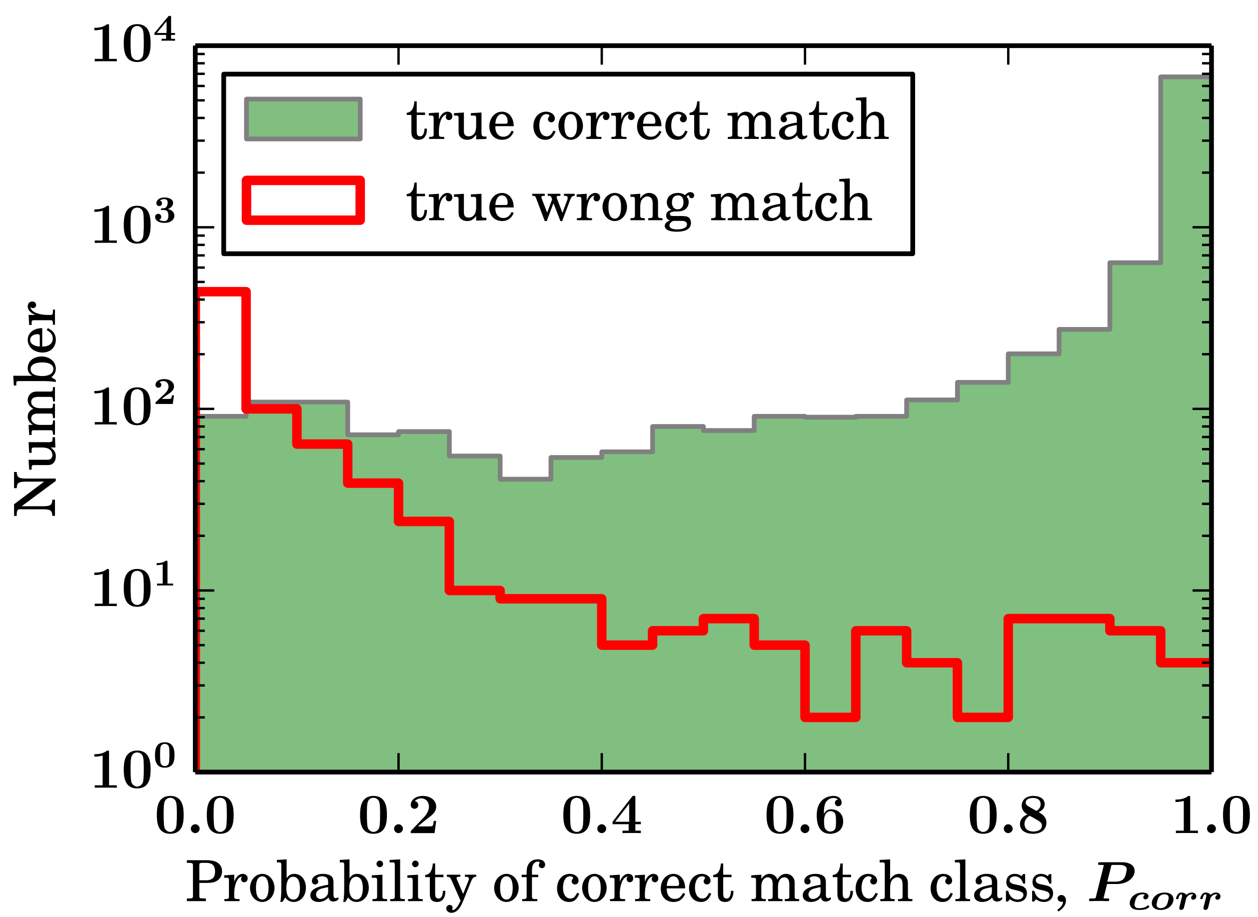
Improvements with Machine Learning

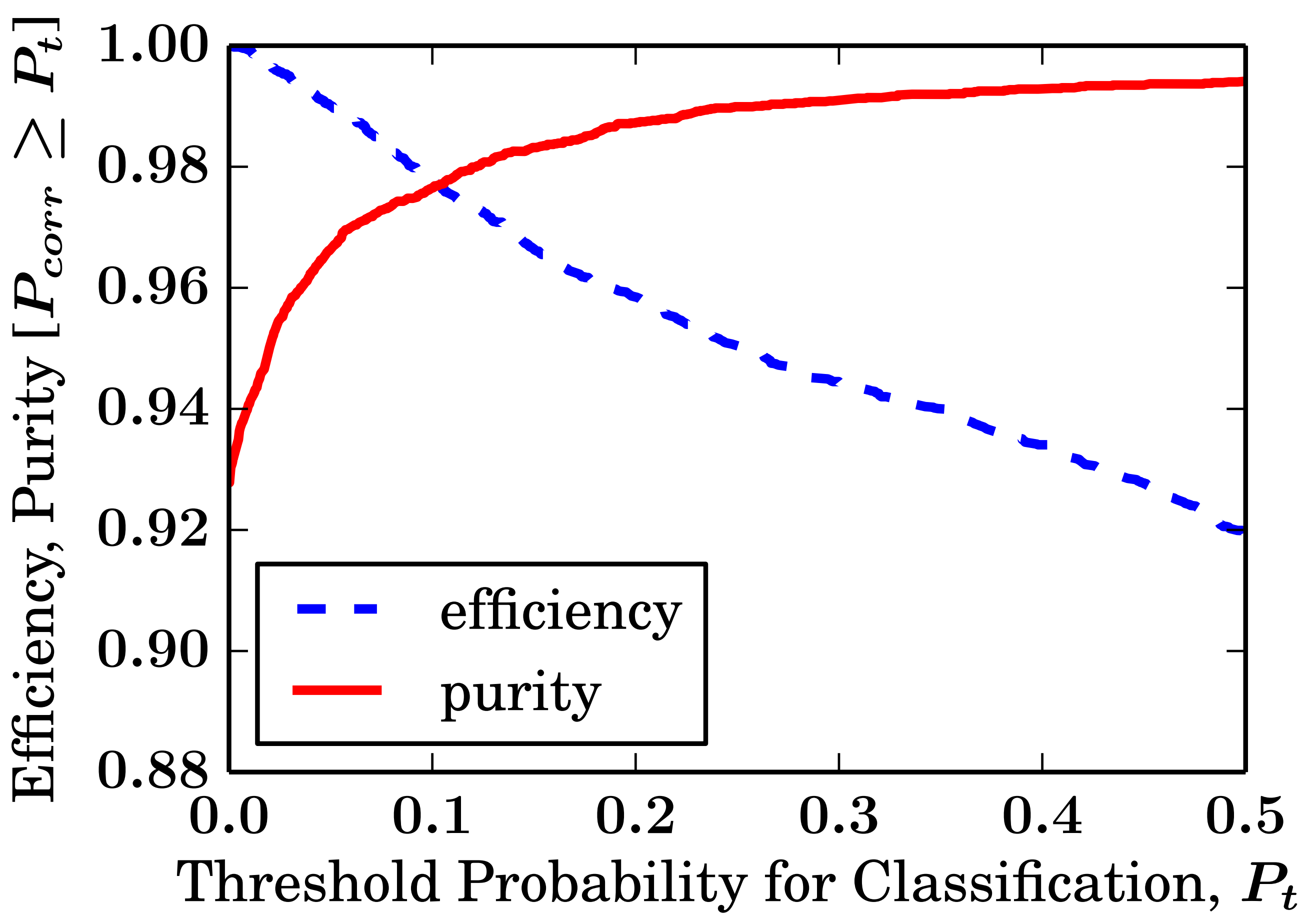
- We would like some way of quantifying the probability of a correct match for each SN, while also improving the matching accuracy, if possible
- After initial DLR matching algorithm, implement Random Forests for binary classification into {correct match, wrong match}
- Features of the SN-matched host pairs are used to train the classifier



Improvements with Machine Learning: Results

- Applying the trained ML classifier to a validation set, we can obtain (for each SN-host match) the probability of a correct match
- Fixing the efficiency at 98%, we find that **ML boosts the matching accuracy (purity) up to ~97%**



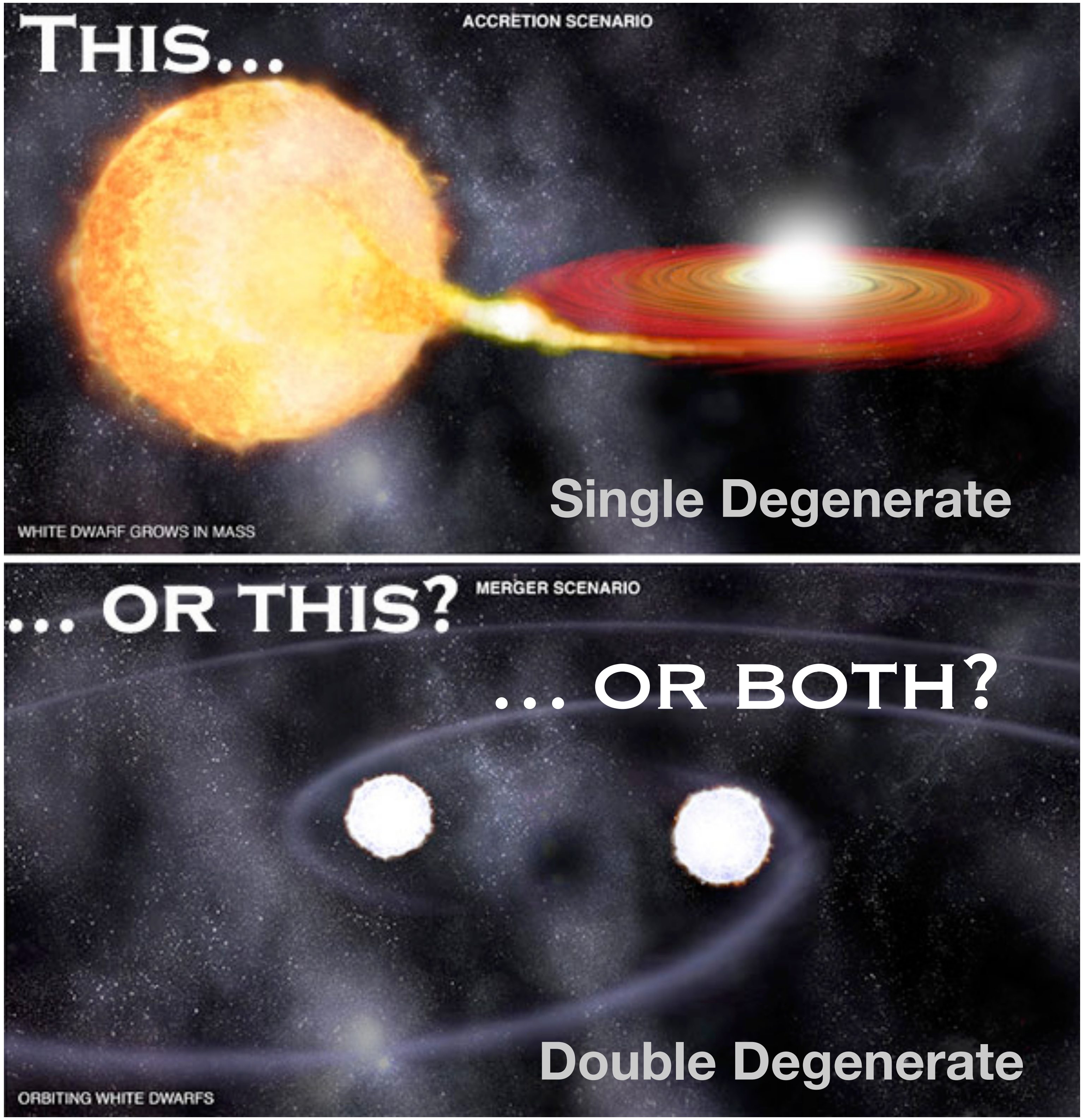


Next Steps

- This paper is headed for DES collaboration-wide review in a couple weeks
- A follow-up paper will focus specifically on SNe Ia and DES, including propagating the effects of host galaxy mismatches to photometric SN classification and biases on cosmological parameters
- Train the ML classifier on real DES galaxy catalogs so we can begin assigning ML probabilities to actual DES SNe

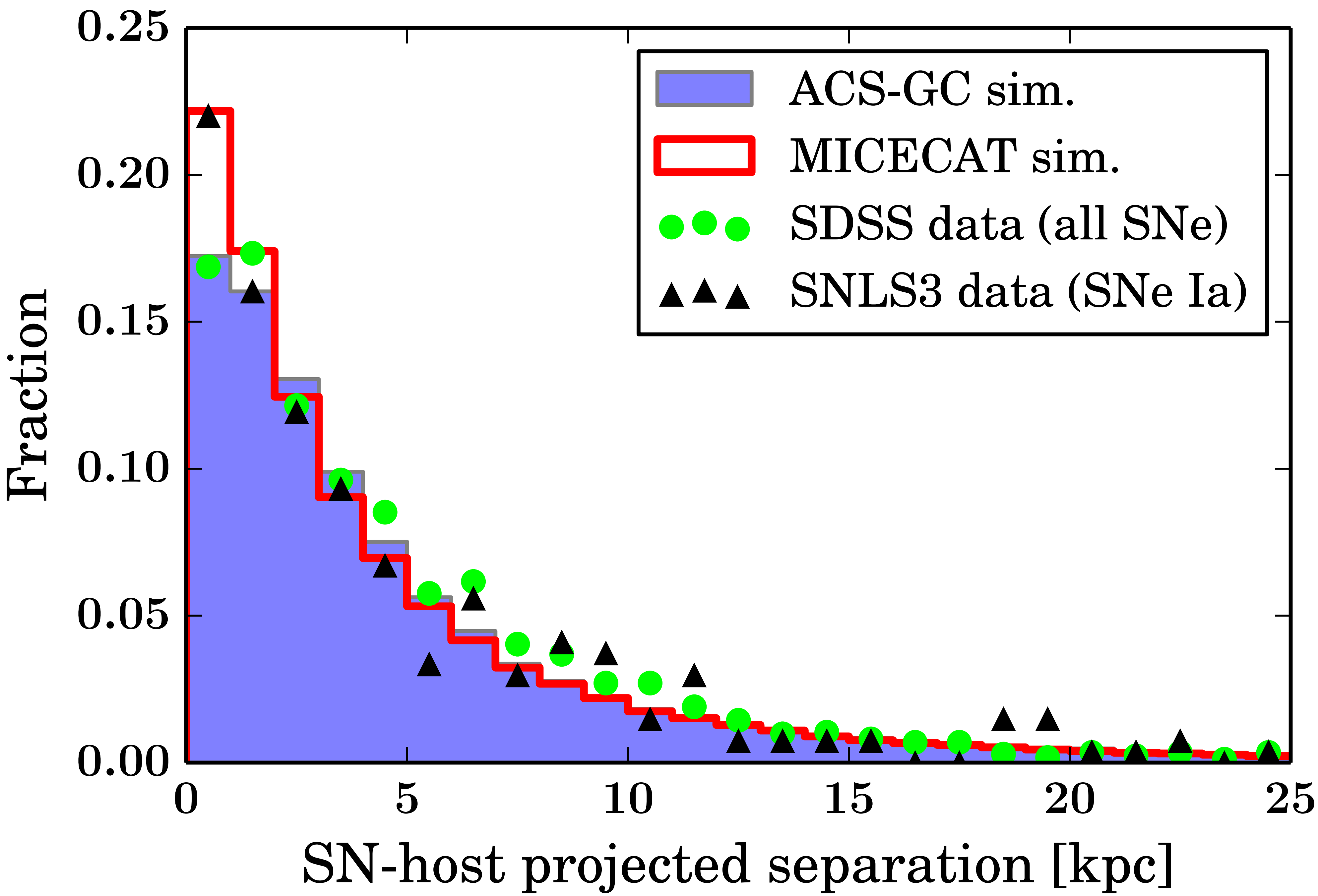
EXTRA SLIDES

Type Ia Supernova Progenitors



NASA/CXC/M.Weiss/Bad Astronomer

Comparing SN Simulations with Data



Machine Learning: Definitions

		TRUE CLASS	
		Correct Match	Wrong Match
PREDICTED CLASS	Correct Match	True Positives T_P	False Positives F_P
	Wrong Match	False Negatives F_N	True Negatives T_N

- Efficiency = fraction of true correct matches recovered by the classifier
  - $T_P / (T_P + F_N)$
- Purity = the accuracy with which objects are classified as correct matches
  - $T_P / (T_P + F_P)$