Supernovae Demography and Rates based on Machine Learning Classification

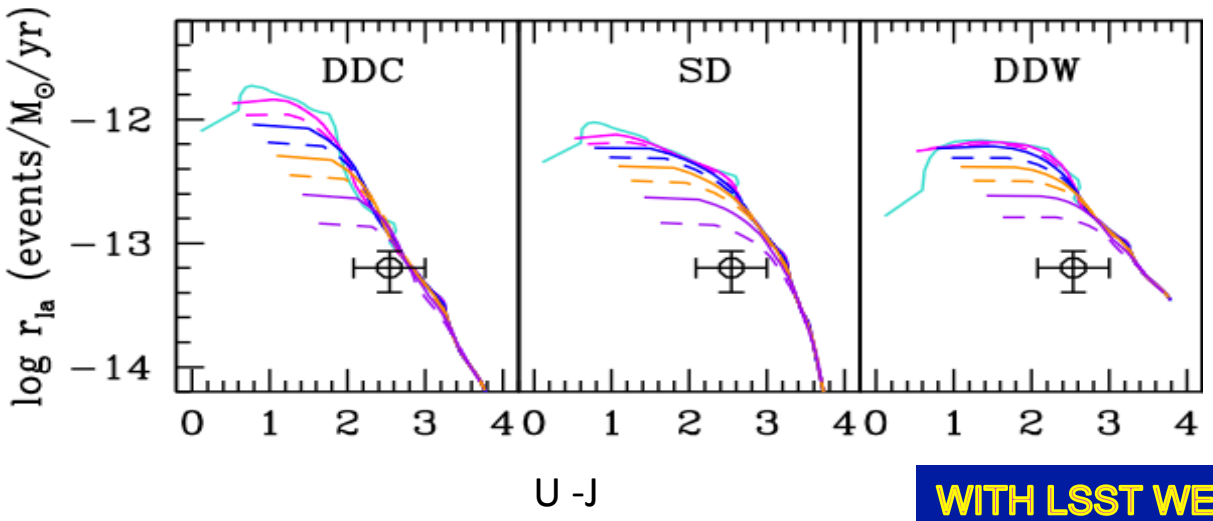
The ultimate goal is to relate the various kinds of SN explosions to their stellar progenitors

A few fundamental questions

- ❖Are SNIa mainly produced in SD or DD systems?
- ❖Do they come from different channels in young and in old stellar populations?
- ❖Which is the mass range of CC SNe progenitors?
- ❖Which are the progenitors of the various subtypes of CC SNe?
- ❖How do the evolutionary scenarios leading to SNe depend on metallicity?

To answer these questions we need a LARGE and UNBIASED sample of SN events in galaxies with well characterized stellar populations (SFH and metallicity)

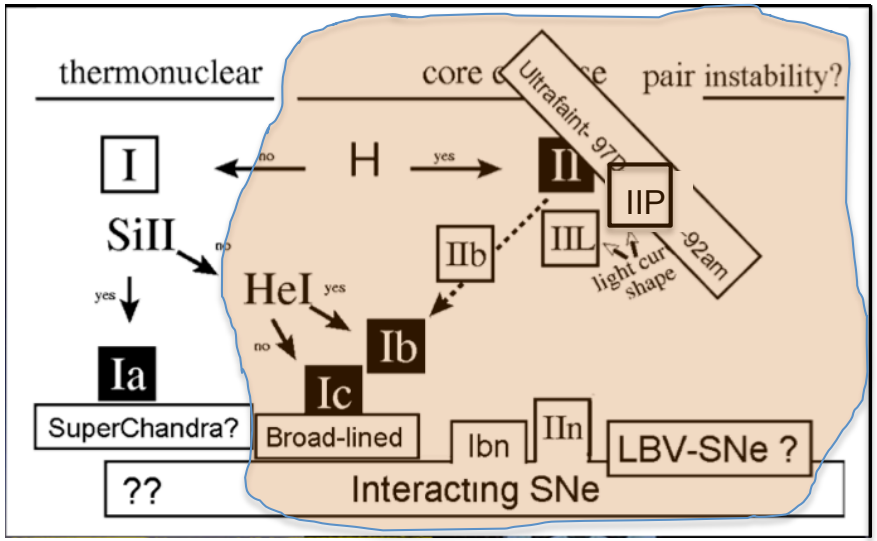
LARGE – for statistical reasons
UNBIASED – to equally sample all evolutionary channels
PARENT GALAXY PROPERTIES – to relate statistically the SN kinds to the age and metallicity of their progenitors



SNIa:
different progenitor models imply a different slope for the scaling of the SNIa rate with the color of the parent galaxy.

The Zoo of CC Supernovae

Credit: Terreran, Pd 2017; Turatto+ 2007



WITH LSST WE AIM AT collecting photometric data on a well defined galaxy sample to determine:

- **redshift**
- **mass and SFH of the galaxies**
- **SNe light curves for their classification**
- **means to evaluate the completeness (→rates)**

To improve on the SUDARE statistics (~ 250 SNe) by a factor of ~ 10 in ~ 3 yrs we target:

- **a large fraction of the whole sky at z~0 to get the SNIa in passive galaxies of various colors**
- **a few 10s of square degrees at z ~0.3-0.5 and**
- **a few sq deg at z>0.5**

to map the dependence of the rates on SFR and z

This database will also enable an unbiased measure of the SN diversity and evolution of the cosmic mix

R&D on data mining and machine learning for classification of variable sources

Feature extraction and selection

Following the experience on CRTS (Catalina Real-time Transient Survey; *Drake et al. 2009*), from light curves (LCs) we extract a parameter space formed by a set of statistical features (*D'Isanto et al. 2015*), useful to evaluate the information entropy carried by each feature and to select the minimal set of parameters representative of the knowledge domain (knowledge base or KB).

Classification of variable objects

Several self-adaptive classifiers, e.g. MLPQNA (*Brescia et al. 2012*) Random Forest (*Breiman 2001*), K-NN (*Hastie et al. 2001*), GNG (*Fritzke 1995*), based on both supervised and unsupervised paradigms of machine learning (ML), are trained on the KB and validated on blind test sets to identify and classify variable objects.

The ability to recognize and quantify the differences among light curves with ML methods requires many instances of light curves for each class of interest. That's why LSST may offer an invaluable chance to exploit such methodologies.

We started to tune and test our methods on the SUDARE data (*Cappellaro et al. 2015*). From ~200 LCs we extracted a 59-D parameter space on which GNG and Random Forest are used to learn the classification of different types of SNs. Results appear promising, even though the extremely limited KB available.

Partecipanti e relative responsabilità/interessi

Laura Greggio (INAF – OAPd) : theory of progenitor's models, Stellar Pop modelling
Maria Teresa Botticella (INAF – OACN) : data analysis, rates' computation, SED fitting
Giuseppe Riccio (INAF - OACN) : SN types classification and redshift

Expertise/Interests of Collaborators:
Filippo Mannucci (INAF – OAArcetri) : SN rates and galaxies' properties
Enrico Cappellaro (INAF- OAPd) : SNe properties and rates
Massimo della Valle (INAF-OACN) : SNe and CNe properties and rates
Massimo Brescia (INAF – OACN): Machine Learning Classification tools
Stefano Cavuoti (Univ. Federico II, Na) : Machine Learning Classification tools
Giuseppe Longo (Univ. Federico II, Na) : Machine Learning Classification tools
Giuseppe Angora (Univ. Federico II, Na) : Machine Learning Classification tools
Maurizio Paolillo (Univ. Federico II, Na) : AGN detection as contaminants of SN sample

ALL: comparison to models and interpretation

Connessioni con osservazioni di altre *facilities*

We anticipate synergy with:

EUCLID – IR data will prove very important for the SED fitting of the sample galaxies
AND to investigate on extinction and its effect on SNe detection

WFIRST - ?

SOXs – spectra of a number of (nearby) SNe will allow validation of the photometric typing and of redshift determination

VST – integration of light curve sampling for a portion of the surveyed area, to assess the quality of the photometric typing and improve it

Tipo di analisi dati prevista/necessaria, inclusa di sviluppo e cadenza temporale

In the next 3 yrs we plan to develop our classification tools and test them with SUDARE data, as well as with simulations

After the first year of operation we will be ready to start the analysis:

-determine the photometric redshifts of the sample galaxies
-measure mass, extinction, rest frame colors of the sample galaxies from SED fitting
These results will be obtained using public tools available in the literature

-photometric typing of the SNe
This will be done both with public tools and with codes specifically developped, trained and tested for our purpose

-assessment of completeness for the various SN kinds
-measurement of the rates of different kinds of SNe (and CNe) in different galaxy types
The software for this task is well in hand and tested.

To be clarified whether some of these measurements will be provided by the LSST consortium catalog.