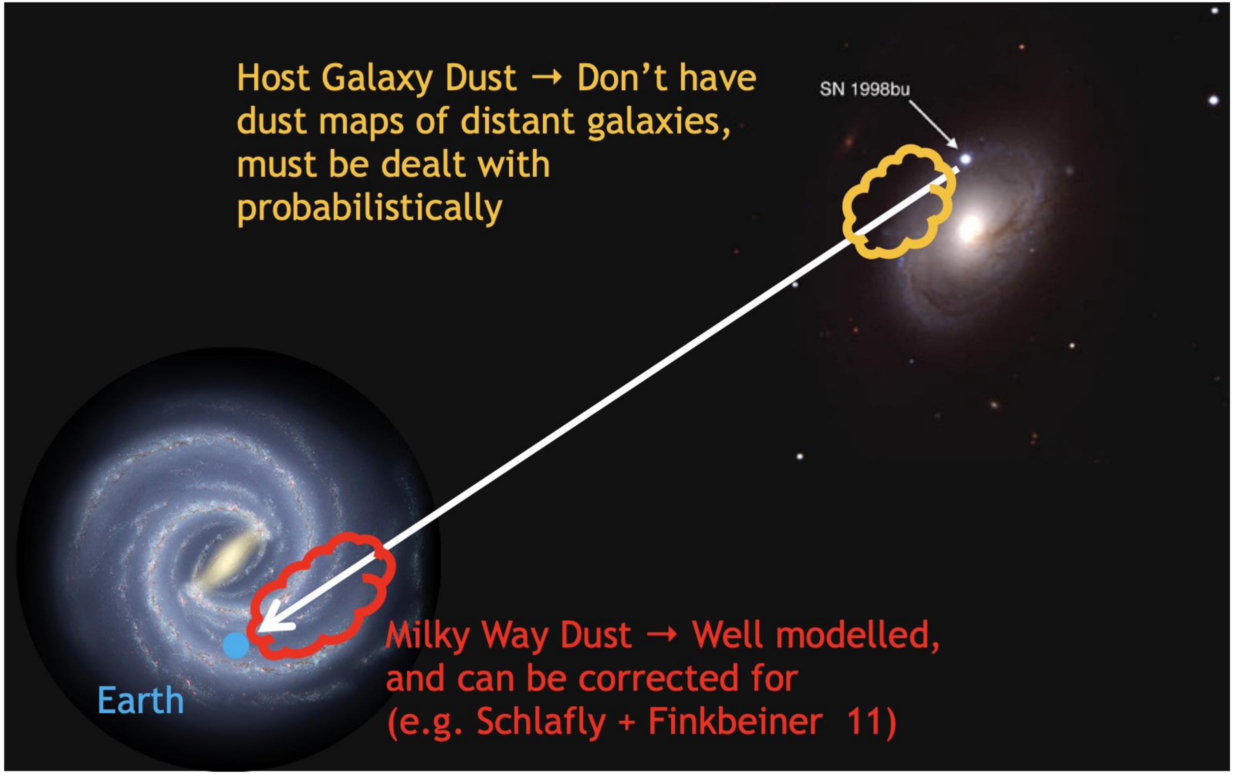
BayeSN: Scalable Hierarchical Modelling of Type Ia Supernovae

Matt Grayling, Kaisey Mandel, Stephen Thorp, Gautham Narayan, Sam Ward

Motivation



NASA/JPL-Caltech/ESO/R. Hurt
Nicholas B. Suntzeff

Dust vs Intrinsic Differences

UNIVERSITY OF CAMBRIDGE

$$\mu^s = m_B^s - M + \alpha x_1^s + \beta c^s$$

- Tripp formula - one parameter for two separate effects
- Correctly handling dust is key for both SN Ia astrophysics and cosmology
- If intrinsic effect misattributed to dust, could lead to bias

Hierarchical Modelling

UNIVERSITY OF CAMBRIDGE

- For typical Bayesian inference, each supernova would be fit separately

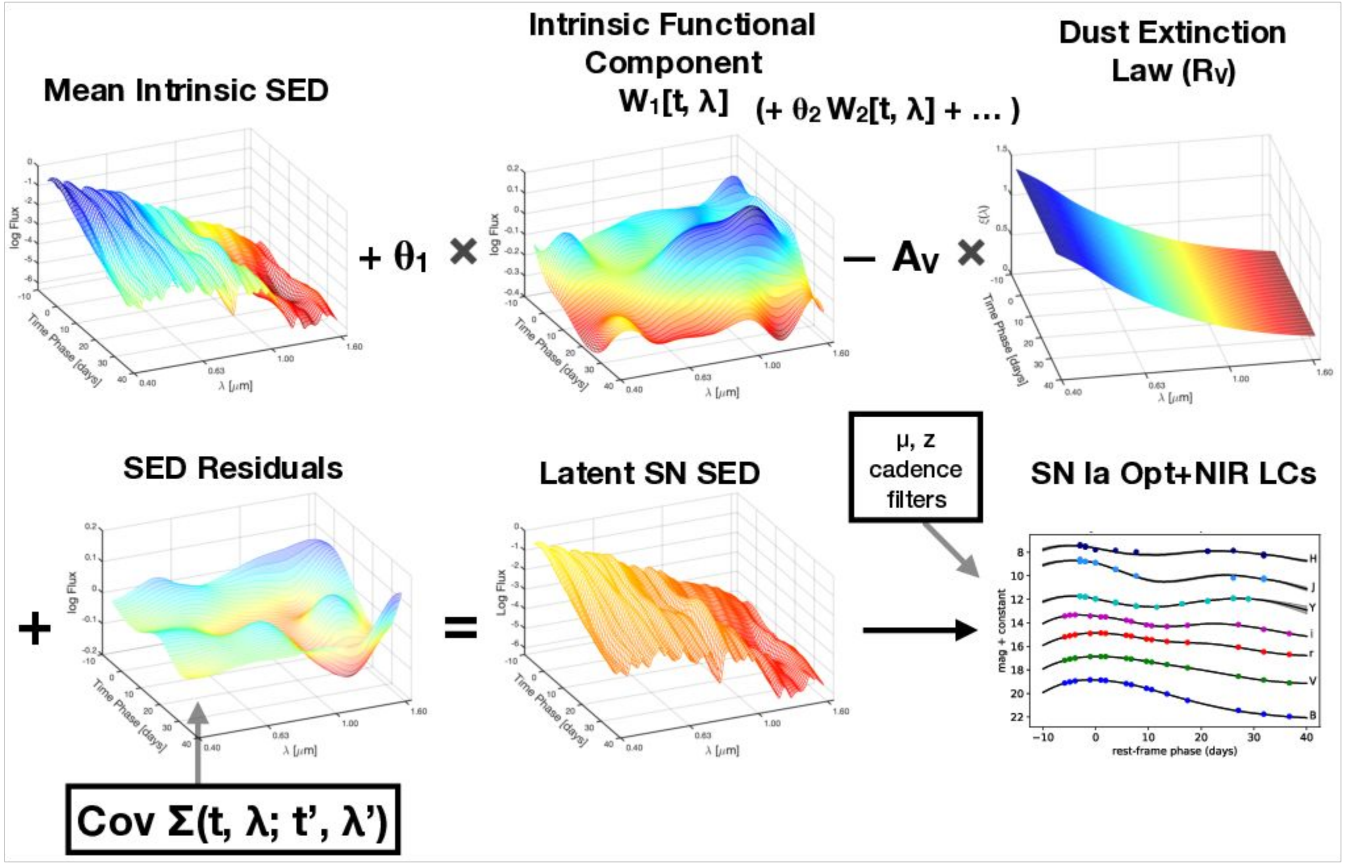
$$R_V \sim U(1,5)$$

- For a hierarchical model, we jointly the sample the posteriors of global and individual supernova properties

$$R_V \sim N(\mu_R, \sigma_R^2)$$

- We can model intrinsic variation and dust properties at the population level as two separate effects

The BayeSN Model



Mandel (2020)

Advantages of BayeSN

UNIVERSITY OF CAMBRIDGE

- Hierarchical Bayesian Model
  - Allows for separate treatment of dust and intrinsic variation
  - Better estimation of global properties
- Full SED model - all colour/magnitude information used when estimating distance rather than just single colour/magnitude pairs
- Model extends into NIR wavelengths - better SN Ia standardisation and constraint of host galaxy dust

Challenges with BayeSN

UNIVERSITY OF CAMBRIDGE

- Lots of parameters (e.g ~400 global, ~4200 latent parameters for Thorp 2021)
- Complex likelihood evaluation:
    - Compute SED model
    - Evaluate numerical integrals through different filters, for each time of observation and each supernova
- Computationally expensive

Scaling BayeSN

UNIVERSITY OF CAMBRIDGE

Problem:

Scaling BayeSN for next generation data sets without compromising functionality

Scaling BayeSN

Problem:

Scaling BayeSN for next generation data sets without compromising functionality

Solution:

pyro-ppl/**numpyro**

Probabilistic programming with NumPy powered by JAX for autograd and JIT compilation to GPU/TPU/CPU.

82 Contributors | 529 Used by | 2k Stars | 179 Forks

Jax and Numpyro

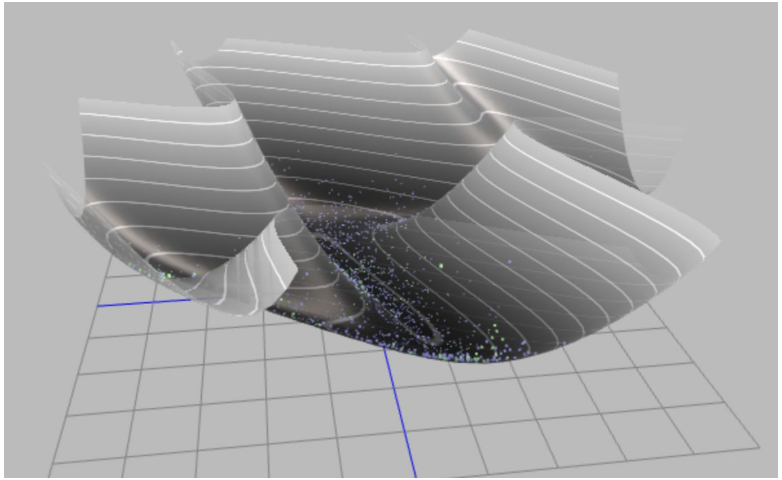
UNIVERSITY OF CAMBRIDGE

Jax

- Numpy with JIT compilation for GPUs
- Supports autodiff (derivatives for free), ideal for Hamiltonian Monte Carlo
- Very efficient matrix/tensor operations

Numpyro

- Probabilistic programming package for Python built on Jax



Credit: Alex Rogozhnikov

Vectorized likelihood evaluation + GPUs = Fast Bayesian inference

Benchmarks

UNIVERSITY OF CAMBRIDGE

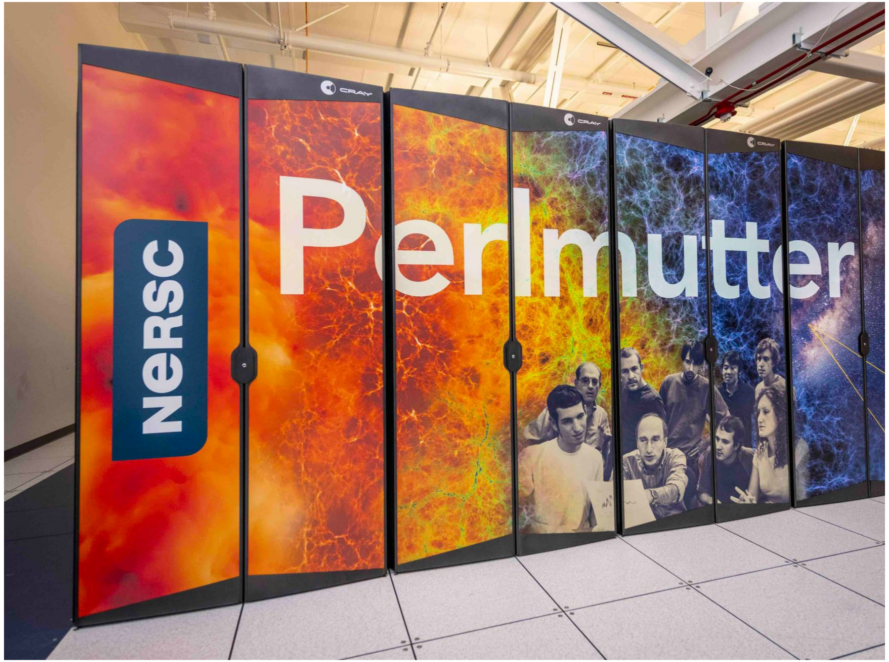
On Perlmutter, using 4 GPUs

Thorp+21:

- 157 *griz* Foundation light curves
- Training ~1 day → ~10 minutes
- Fits ~2 minutes per object → ~4 minutes for all 157

Mandel+20:

- 79 optical+NIR light curves
- Training ~5 days → ~20 minutes
- Fits ~40 minutes per object → ~18 minutes for all 79



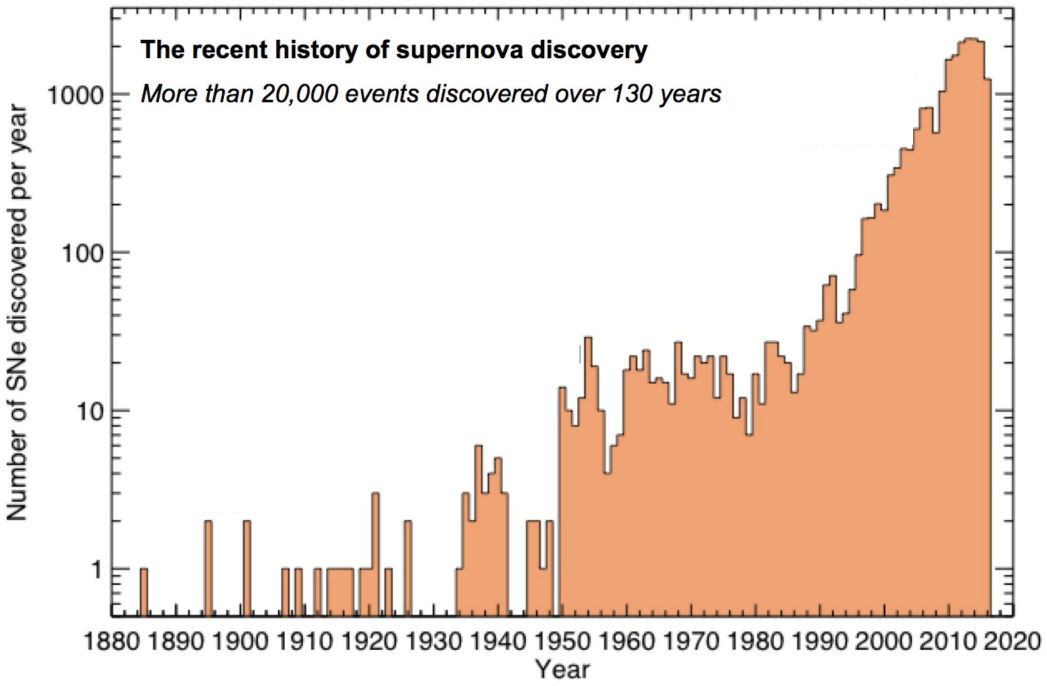
Credit: NERSC

Scalability

Testing scalability with simulated Foundation-like SNe:

- 100 SNe
  - Training - 9 mins
  - Fitting - 3 mins
- 1000 SNe
  - Training - 21 mins
  - Fitting - 18 mins

Efficiency of tensor operations improves as sample size grows, model is scalable by construction

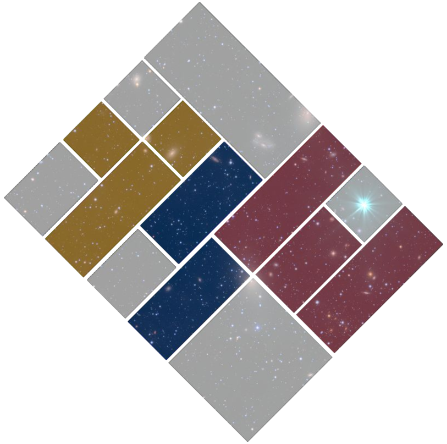


Credit: Mark Sullivan

Applying the Model

UNIVERSITY OF CAMBRIDGE

- Previously, training and fitting on data-sets > 10,000 SNe would have been computationally unfeasible - now achievable in relatively short timescales
- Work ongoing to implement BayeSN within SNANA for cosmological analyses
- Planned hierarchical dust analysis of hundreds of SNe Ia from the Young Supernova Experiment (YSE), as well as archival samples
- Ideal framework to look at dust and intrinsic distributions as a function of environment



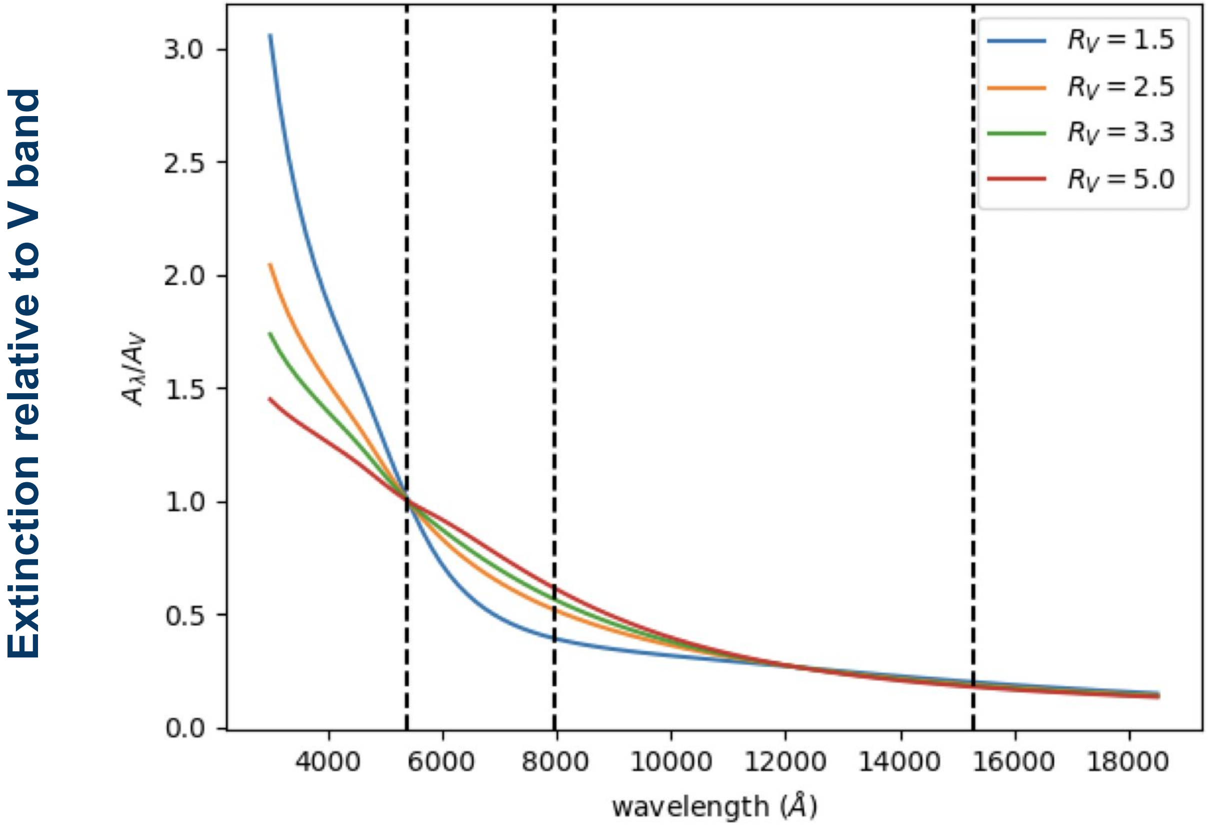
Conclusions

UNIVERSITY OF CAMBRIDGE

- BayeSN is a hierarchical model for fitting type Ia supernovae with several advantages over current approaches
- Through the use of jax/numpyro and GPU acceleration, we have been able to achieve ~100x performance speed ups, making the use of BayeSN feasible for large surveys
- BayeSN being implemented in SNANA for cosmological analysis
- Lots of potential for analysis of dust and intrinsic distributions of SNe Ia

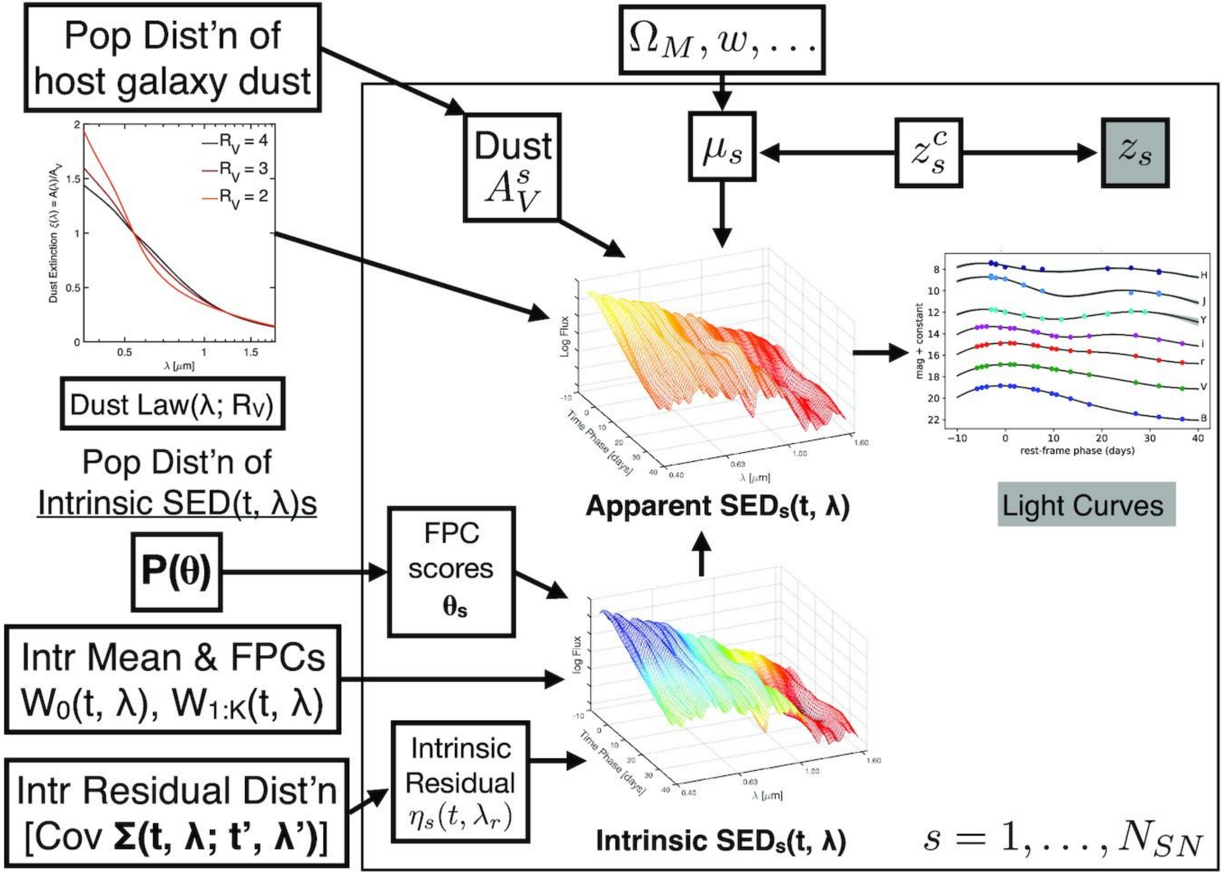
Dust

Extinction vs. wavelength



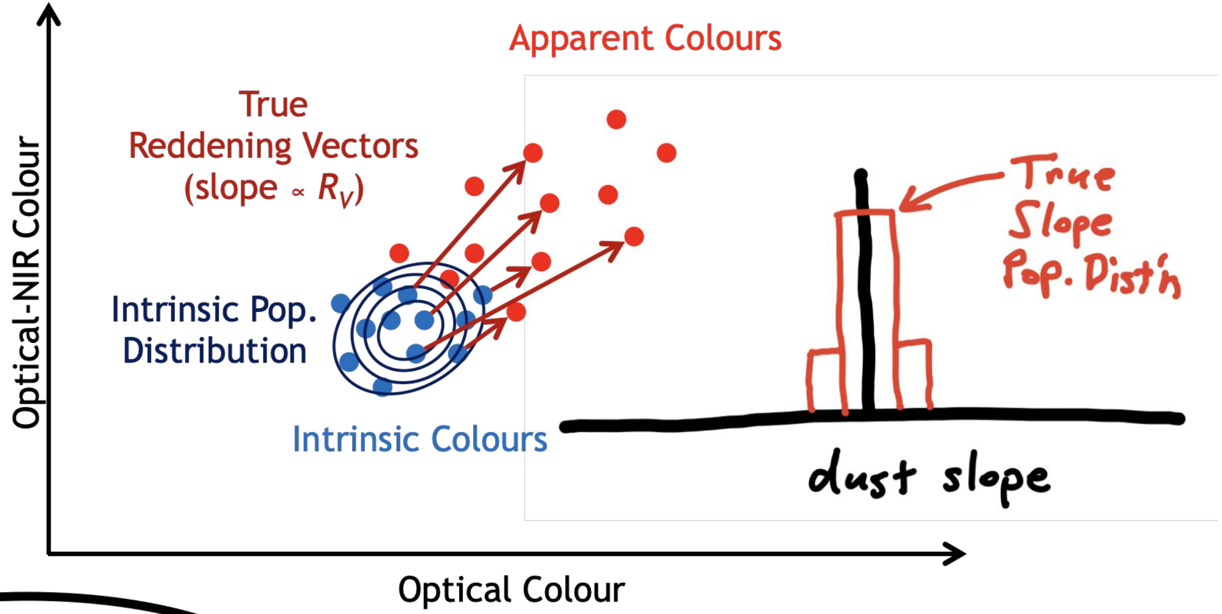
Lower Rv → steeper dust law

The BayeSN Model



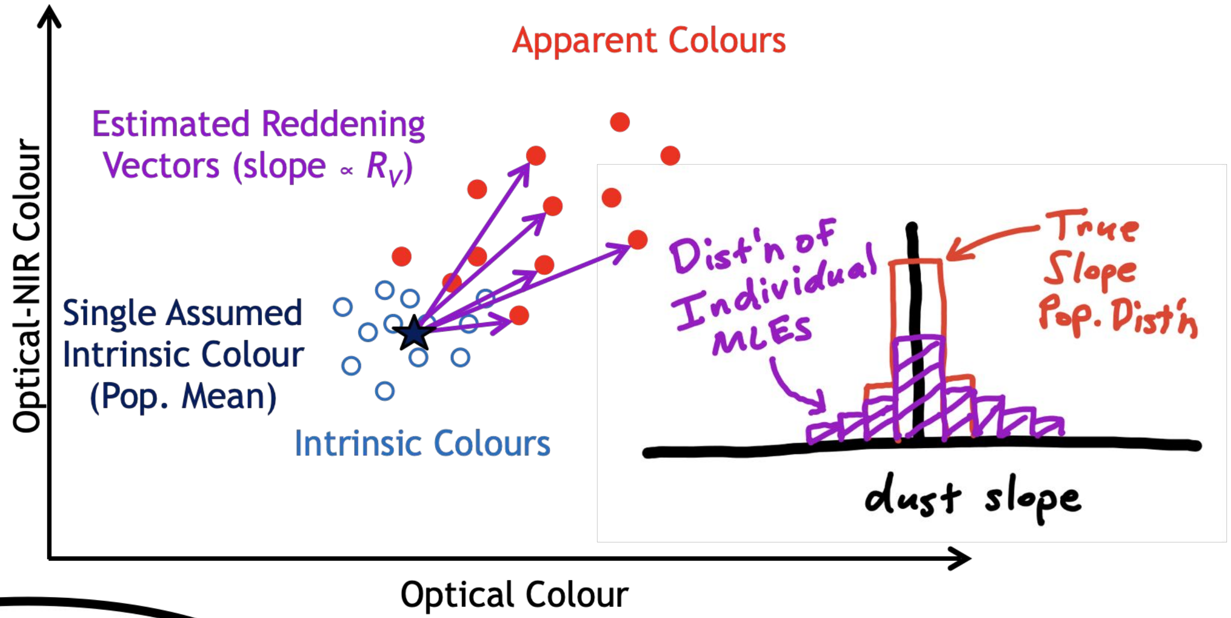
Mandel (2020)

Why include intrinsic variation?



Need to account for intrinsic colour scatter when inferring R_V!

Why include intrinsic variation?

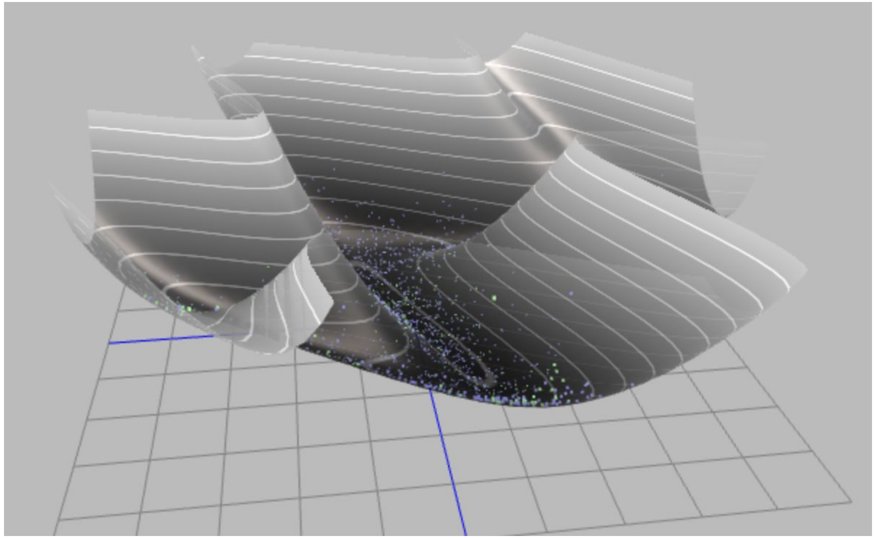


Otherwise width of R_V distribution will be overestimated!

Using BayeSN

UNIVERSITY OF CAMBRIDGE

- Two modes of use
  - Training - Conditioning global parameter values on data
  - Fitting - Inference of supernova properties with fixed global parameters
- Uses Hamiltonian Monte Carlo (HMC) to sample posteriors
  - MCMC method using gradient to make efficient steps



Credit: Alex Rogozhnikov

Application to Supernova Photometry

UNIVERSITY OF CAMBRIDGE

- Light curves are not of a fixed shape, but we want vectorized calculations
- For a given data set, all light curves are padded with zeros to match the longest time series. Numpyro supports masking - padded data points do not contribute to likelihood
- Allows all flux integrals across all SNe, phases and bands to be calculated as a single tensor operation - calculations scale very efficiently
- Using HMC, we can fit multiple SNe simultaneously with shared flux integral calculations