

# LUMINA: GPU-Accelerated Monte Carlo Radiative Transfer and Machine Learning Spectral Inference for Type Ia Supernovae

Kyeong Jae Han<sup>1\*</sup>

<sup>1</sup>Department of Physics and Astronomy, Seoul National University, Seoul 08826, Republic of Korea

Accepted XXX. Received XXX; in original form XXX

## ABSTRACT

We present LUMINA, a GPU-accelerated one-dimensional Monte Carlo radiative transfer code for spectral synthesis of Type Ia supernovae (SNe Ia), and LUMINA-ML, a companion machine learning pipeline for Bayesian parameter inference from observed spectra. LUMINA is implemented in C/CUDA and faithfully reproduces the physics of the established TARDIS code, including Sobolev line transfer, macro-atom fluorescence, and non-local thermodynamic equilibrium (NLTE) rate equations for six ion pairs (Si, Ca, Fe, S, Co, Ni II/III; 3547 levels, 64 322 lines). The GPU implementation achieves a factor of  $\sim 10$  speedup over the equivalent CPU code, enabling the generation of large training sets for machine learning applications. LUMINA-ML uses a multi-stage recursive Bayesian framework: a neural network emulator (five-layer MLP) is trained on  $\sim 25\,000$  LUMINA simulations, then used as a fast surrogate within Markov Chain Monte Carlo (MCMC), nested sampling, and simulation-based inference (SBI) to derive posterior distributions over 15–63 physical and compositional parameters. Applied to SN 2011fe at  $B$ -maximum, the Stage 1 (15-dimensional) MCMC posterior yields  $\log L = 42.80 \pm 0.01 \text{ erg s}^{-1}$ ,  $v_{\text{inner}} = 7553 \pm 25 \text{ km s}^{-1}$ , and  $t_{\text{exp}} = 10.1 \pm 0.1 \text{ d}$ . The Stage 2 analysis extends the parameter space to 63 dimensions by resolving the ejecta into six radial composition zones with eight free elemental abundances each, enabling spatially resolved abundance tomography. The emulator achieves median spectral RMS residuals of 0.36–0.39 per cent in the optical, and the full pipeline — from simulation to posterior — runs in hours on a single GPU, compared to weeks with direct forward modelling.

**Key words:** radiative transfer – methods: numerical – methods: statistical – supernovae: general – supernovae: individual: SN 2011fe

## 1 INTRODUCTION

Type Ia supernovae (SNe Ia) are thermonuclear explosions of carbon–oxygen white dwarfs that serve as standardizable cosmological distance indicators (Phillips 1993; Riess et al. 1998; Perlmutter et al. 1999). Detailed spectroscopic analysis of SNe Ia provides critical constraints on the explosion physics, progenitor systems, and nucleosynthetic yields (Nomoto et al. 1984; Branch et al. 2006). However, extracting quantitative physical parameters from SNe Ia spectra requires computationally expensive radiative transfer simulations that couple photon transport with atomic physics in rapidly expanding ejecta.

Monte Carlo radiative transfer (MCRT) codes such as TARDIS (Kerzendorf & Sim 2014), SEDONA (Kasen et al. 2006), and ARTIS (Sim 2007) have become essential tools for SNe Ia spectral modelling. Among these, TARDIS has emerged as a widely used, open-source framework for one-dimensional spectral synthesis under the Sobolev approximation (Sobolev 1960). While physically rigorous, TARDIS is implemented in Python/Cython and requires  $\sim$ minutes per model evaluation, making systematic parameter exploration across high-dimensional spaces prohibitively expensive.

Recent efforts have sought to accelerate spectral inference through machine learning surrogates. Kerzendorf et al. (2021) introduced

DALEK, a deep-learning emulator for TARDIS, while O’Brien et al. (2021) and Vogl et al. (2020) demonstrated neural network emulators for supernova spectral synthesis. These approaches replace the full forward model with a fast approximation, enabling Bayesian inference over physically meaningful parameter spaces.

In this paper, we present two complementary tools:

(i) **LUMINA**: a C/CUDA reimplementation of the core TARDIS physics that achieves  $\sim 10\times$  speedup through GPU parallelization, with extensions including NLTE rate equations for six ion pairs, charge exchange coupling, and dynamic macro-atom transition probabilities.

(ii) **LUMINA-ML**: a multi-stage Bayesian inference pipeline that trains neural network emulators on LUMINA simulations and performs posterior estimation via MCMC, nested sampling, and simulation-based inference (SBI).

The structure of this paper is as follows. Section 2 describes the LUMINA radiative transfer code, including the physics implementation, NLTE solver, and GPU architecture. Section 3 presents the LUMINA-ML pipeline, covering the emulator architecture, training strategy, and inference methods. Section 4 applies the full framework to SN 2011fe at  $B$ -maximum, presenting posterior parameter estimates and spectral fits. Section 5 discusses the results and future directions. Section 6 summarizes our conclusions.

\* E-mail: kjhan@snu.ac.kr

## 2 THE LUMINA RADIATIVE TRANSFER CODE

LUMINA is a one-dimensional Monte Carlo radiative transfer code designed for rapid spectral synthesis of SNe Ia. It is implemented in C (CPU) and CUDA (GPU), comprising approximately 3 900 lines of source code across six modules. The code faithfully reproduces the physics of TARDIS (Kerzendorf & Sim 2014), enabling direct validation against the established Python/Cython implementation.

### 2.1 Ejecta model

The supernova ejecta are modelled as a spherically symmetric, homologously expanding envelope ( $v \propto r/t$ ) divided into  $N_{\text{shell}} = 30$  radial shells. The velocity grid spans from  $v_{\text{inner}}$  (the inner boundary/photosphere) to  $v_{\text{outer}} = 25\,000 \text{ km s}^{-1}$  with uniform spacing  $\Delta v = (v_{\text{outer}} - v_{\text{inner}})/N_{\text{shell}}$ .

The density profile follows a broken power law:

$$\rho(v, t) = \rho_0 \left( \frac{v}{v_{\text{inner}}} \right)^n \left( \frac{t_{\text{ref}}}{t_{\text{exp}}} \right)^3, \quad (1)$$

where  $n = n_{\text{inner}}$  for  $v < v_{\text{break}}$  and  $n = n_{\text{outer}}$  for  $v \geq v_{\text{break}}$ , and  $t_{\text{ref}} = 19 \text{ d}$  is the reference epoch. The time scaling factor  $(t_{\text{ref}}/t_{\text{exp}})^3$  accounts for homologous dilution.

The composition is specified per shell, with 11 elements tracked: C, O, Mg, Si, S, Ca, Ti, Cr, Fe, Co, and Ni. The Ni<sup>56</sup> decay chain (Ni<sup>56</sup> → Co<sup>56</sup> → Fe<sup>56</sup>) is computed via Bateman equations at epoch  $t_{\text{exp}}$ , with decay constants  $\lambda_{\text{Ni}} = 0.0788 \text{ d}^{-1}$  and  $\lambda_{\text{Co}} = 0.00622 \text{ d}^{-1}$ . Oxygen serves as the filler element (mass fraction  $X_{\text{O}} = 1 - \sum X_i$ ).

### 2.2 Photon transport

Photon packets are initialized at the inner boundary with a blackbody spectrum at temperature  $T_{\text{inner}}$  and total luminosity  $L_{\text{inner}}$ . Each packet carries equal energy  $\epsilon = L_{\text{inner}}/N_{\text{packets}}$ .

Packet propagation employs the Sobolev approximation (Sobolev 1960; Lucy 1999): in the comoving frame, each spectral line appears as a delta function in frequency, and the interaction probability is determined by the Sobolev optical depth:

$$\tau_{\text{Sob}} = \frac{\pi e^2}{m_e c} f_{lu} \lambda_{lu} t_{\text{exp}} n_l \left( 1 - \frac{g_l n_u}{g_u n_l} \right), \quad (2)$$

where  $f_{lu}$  is the oscillator strength,  $\lambda_{lu}$  the rest wavelength, and  $n_l$ ,  $n_u$  are the lower and upper level populations.

As a packet traverses a shell, its comoving-frame frequency shifts monotonically due to the velocity gradient. We perform a *full Doppler sweep* from the entry frequency  $\nu_{\text{entry}}$  to the exit frequency  $\nu_{\text{exit}}$ , accumulating the optical depths of all intervening lines. The interaction site is selected by sampling against the cumulative optical depth. This approach captures all line interactions, unlike a narrow frequency window which can miss lines in optically thick regions.

### 2.3 Line interaction: macro-atom treatment

Upon line interaction, a photon packet activates the corresponding atomic level and undergoes the macro-atom cascade (Lucy 2002, 2003). The activated level can:

- (i) *Emit* a photon (radiative de-excitation), with probability  $\propto A_{ul}\beta(\tau)$  where  $\beta(\tau) = (1 - e^{-\tau})/\tau$  is the Sobolev escape probability;

- (ii) *Transition internally downward*, with probability  $\propto A_{ul}[1 - \beta(\tau)]$ ;

- (iii) *Transition internally upward*, with probability  $\propto B_{lu} W B_{\nu}(T_{\text{rad}}, \nu)$ , where  $W$  is the dilution factor.

The code tracks 10 791 energy levels and 411 756 macro-atom transitions across all included species. A downbranching/fluorescence channel (Lucy 2002) is also implemented, with wavelength-dependent thermalization fractions:  $\varepsilon = 0.80$  (infrared), 0.35 (optical), and 0.175 (ultraviolet).

### 2.4 Plasma state and convergence

The radiation field is characterized by two quantities per shell: the dilution factor  $W$  and the radiation temperature  $T_{\text{rad}}$ , derived from the Monte Carlo estimators for mean intensity  $J$  and frequency-weighted mean  $\bar{\nu}$ :

$$W = \frac{J}{4\sigma T_{\text{rad}}^4 t_{\text{sim}} V_{\text{shell}}}, \quad T_{\text{rad}} = \frac{h\bar{\nu}}{3.832 k_B}. \quad (3)$$

Ionization balance follows the nebular approximation with dilution:

$$\frac{n_{i+1} n_e}{n_i} = W \Phi(T_{\text{rad}}) \zeta(T_{\text{rad}}), \quad (4)$$

where  $\Phi$  is the Saha factor computed at  $T_{\text{rad}}$  and  $\zeta$  is a tabulated departure coefficient. Partition functions are evaluated at  $T_{\text{rad}}$  via Boltzmann summation over all levels of each ion.

The inner boundary temperature is updated iteratively via luminosity feedback:

$$T_{\text{inner}}^{(n+1)} = T_{\text{inner}}^{(n)} \left( \frac{L_{\text{emitted}}}{L_{\text{requested}}} \right)^{-1/2}, \quad (5)$$

with 50 per cent damping applied to  $W$ ,  $T_{\text{rad}}$ , and  $T_{\text{inner}}$  updates. The first three iterations are held frozen to allow the radiation field to stabilize.

After convergence (typically 10–20 iterations with  $2 \times 10^5$  packets), the code achieves agreement with TARDIS to within 0.11 per cent in  $T_{\text{inner}}$ , 0.67 per cent in  $W$ , and 0.28 per cent in  $T_{\text{rad}}$  (Table 1).

### 2.5 NLTE rate equations

For improved accuracy of key diagnostic lines, LUMINA solves full NLTE statistical equilibrium equations for six ion pairs: Si II/III, Ca II/III, Fe II/III, S II/III, Co II/III, and Ni II/III, encompassing 3 547 energy levels and 64 322 NLTE lines.

The rate matrix includes:

- *Radiative bound–bound*: Einstein  $A_{ul}$  (spontaneous emission),  $B_{lu}$  (absorption), and  $B_{ul}$  (stimulated emission), with the radiation field  $J_{\nu}$  estimated from a 1000-bin logarithmic histogram accumulated during Monte Carlo transport.

- *Collisional bound–bound*: van Regemorter formula (van Regemorter 1962) for permitted transitions, with detailed balance for de-excitation.

- *Photoionization*: Kramers cross-sections with threshold frequencies from the atomic database.

- *Radiative recombination*: Milne relation (detailed balance with photoionization).

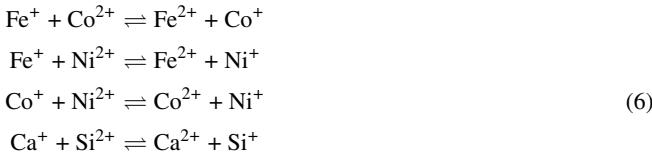
The resulting  $N \times N$  rate matrices (up to  $N = 1362$  for Fe) are solved via LU decomposition. On the GPU, we employ cuBLAS batched LU factorization to solve all 30 shells in parallel, requiring approximately 425 MB of GPU memory.

### 2.5.1 Deferred NLTE

We find that applying NLTE from the first iteration can cause the radiation field to diverge before the LTE solution has converged. A *deferred NLTE* strategy — running the first  $N_{\text{start}}$  iterations in LTE, then activating NLTE for the remaining iterations — produces significantly better results. With  $N_{\text{start}} = 5$  (out of 10 total iterations), the Si II  $\lambda 6355$  absorption trough velocity matches TARDIS to within  $74 \text{ km s}^{-1}$  ( $v_{\text{LUMINA}} = 11\,505 \text{ km s}^{-1}$  vs  $v_{\text{TARDIS}} = 11\,579 \text{ km s}^{-1}$ ), compared to a  $846 \text{ km s}^{-1}$  offset with full NLTE from iteration 0 (Fig. 1).

### 2.5.2 Charge exchange coupling

Four exothermic charge exchange reactions are included to couple the iron-peak ions:



with forward rates of  $10^{-9} \text{ cm}^3 \text{ s}^{-1}$  and reverse rates computed via detailed balance. An iterative wrapper (maximum 5 iterations, 50 per cent damping) ensures convergence of total ion fractions to within 1 per cent.

## 2.6 GPU implementation

The CUDA implementation parallelizes the Monte Carlo transport by assigning one photon packet per GPU thread, with 256 threads per block. Key design choices include:

- *Random number generation*: xoshiro256\*\* (Blackman & Vigna 2021) seeded per thread, avoiding the overhead of cuRAND.
- *Estimator accumulation*: for  $J$  and  $\bar{v}$  estimators, enabling lock-free concurrent updates.
- *NLTE solver*: cuBLAS batched LU factorization for all shells simultaneously.
- *Virtual packets*: On-device virtual packet tracing for low-noise spectrum synthesis.

On an NVIDIA RTX 5000 Ada (CUDA 13.0), LUMINA processes  $2 \times 10^5$  packets in  $\sim 0.7$  s per iteration, achieving a  $\sim 10\times$  speedup over the OpenMP-parallelized CPU version (Table 2). A full production run of  $2 \times 10^6$  packets over 20 iterations completes in  $\sim 14$  s.

## 3 THE LUMINA-ML INFERENCE PIPELINE

LUMINA-ML is a multi-stage Bayesian spectral inference pipeline that leverages LUMINA as a fast forward model and neural network emulators as surrogates for posterior estimation.

### 3.1 Multi-stage parameter space

The inference proceeds in stages of increasing dimensionality:

**Stage 1 (15D):** Global physical parameters — luminosity ( $\log L$ ), photospheric velocity ( $v_{\text{inner}}$ ), reference density ( $\log \rho_0$ ), density exponents ( $n_{\text{inner}}, n_{\text{outer}}$ ), electron temperature ratio ( $T_e/T_{\text{rad}}$ ), zone boundary velocities ( $v_{\text{core}}, v_{\text{wall}}, v_{\text{break}}$ ), time since explosion ( $t_{\text{exp}}$ ), and key abundances ( $X_{\text{Fe},\text{core}}, X_{\text{Si},\text{wall}}, X_{\text{Fe},\text{wall}}, X_{\text{Ni}}, X_{\text{Fe},\text{outer}}$ ).

**Stage 2 (63D):** The 15 physical parameters (relaxed  $\pm 10$  per cent

around Stage 1 best-fit) plus 48 composition parameters: six radial zones  $\times$  eight free elemental abundances (Fe, Si, S, Ca, Ni, Mg, Ti, Cr) per zone.

**Stage 2.5 (63D, refined):** The same 63 parameters, but with prior ranges narrowed based on the Stage 2 SBI posterior (1–99th percentiles plus 30 per cent margin), achieving a volume compression of  $\sim 10^{13.2}$  in parameter space.

### 3.2 Training data generation

Training samples are drawn from the prior via Latin Hypercube Sampling (LHS), ensuring uniform marginal coverage. Physical validity constraints (positive oxygen filler fraction, non-overlapping velocity zones, abundance sum  $\leq 1$ ) are enforced by rejection sampling. For Stage 1, we generate 25 000 models; for Stage 2, 5 000 models.

Each model is executed by LUMINA with  $2 \times 10^5$  packets over 10 iterations, with NLTE enabled from iteration 5 onwards ( $N_{\text{start}} = 5$ ). Multi-device execution (GPU + CPU workers pulling from a shared queue) enables automatic load balancing. The full Stage 1 dataset requires  $\sim 21$  hours on one NVIDIA GPU plus one CPU worker.

### 3.3 Spectral preprocessing

Raw LUMINA spectra (2000 bins, 500–20 000 Å) are preprocessed through the following pipeline:

- (i) *Interpolation* onto a uniform 5 Å grid from 2000 to 10 000 Å (1101 bins).
- (ii) *Adaptive Savitzky–Golay smoothing* with region-dependent windows: 155 Å (UV, 2000–3500 Å), 105 Å (NUV, 3500–4500 Å), 55 Å (optical, 4500–7500 Å), and 105 Å (NIR, 7500–10 000 Å).
- (iii) *Peak normalization* to the maximum flux in the 4000–7000 Å range.
- (iv) *Inverse hyperbolic sine (asinh) transform*:  $y = \text{asinh}(F/F_{\text{ref}})$  with  $F_{\text{ref}} = 0.05$ , equalizing the dynamic range between UV ( $F \sim 0.01\text{--}0.1$ ) and optical ( $F \sim 0.5\text{--}1.0$ ) features.
- (v) *Principal Component Analysis (PCA)*: retaining 99.9 per cent of variance, reducing the spectral dimension from 1101 to  $\sim 289$  (Stage 1) or  $\sim 300$  (Stage 2) components.

### 3.4 Neural network emulator

The emulator is a multi-layer perceptron (MLP) mapping normalized input parameters  $\theta \in \mathbb{R}^D$  to standardized PCA coefficients  $\mathbf{c} \in \mathbb{R}^K$ :

$$\mathbf{c} = f_{\text{MLP}}(\theta; \mathbf{w}). \quad (7)$$

The architecture consists of five hidden layers with dimensions [512, 512, 256, 256, 128], SiLU activation functions, and 1 per cent dropout. The total parameter count is  $\sim 890$  k (Stage 1) or  $\sim 2.2$  M (Stage 2).

Training uses AdamW optimization with learning rate  $5 \times 10^{-4}$ , weight decay  $10^{-5}$ , and cosine annealing with warm restarts ( $T_0 = 100$ ,  $T_{\text{mult}} = 2$ ). The composite loss function combines PCA reconstruction error with feature-specific terms:

$$\mathcal{L} = \mathcal{L}_{\text{PCA}} + \lambda_{\text{feat}} \mathcal{L}_{\text{feat}} + \lambda_{\text{UV}} \mathcal{L}_{\text{UV}}, \quad (8)$$

where  $\mathcal{L}_{\text{feat}}$  weights errors in key spectral windows (Si II  $\lambda 6355$ , Si II W, Ca II H&K, Fe blends, Ca II IR triplet, O I  $\lambda 7774$ ) and  $\mathcal{L}_{\text{UV}}$  penalizes UV/optical ratio errors. We adopt  $\lambda_{\text{feat}} = 0.5$  and  $\lambda_{\text{UV}} = 0.3$ .

Early stopping with patience of 300 epochs terminates training at  $\sim 374$  epochs (Stage 1) and  $\sim 407$  epochs (Stage 2). The resulting emulators achieve median spectral RMS residuals of 0.36 per cent (Stage 1) and 0.39 per cent (Stage 2) on the validation set (Fig. 3).

**Table 1.** Convergence of LUMINA vs TARDIS for the SN 2011fe reference model ( $2 \times 10^5$  packets, 20 iterations).

Quantity	Value	Max. relative error
$T_{\text{inner}}$ (K)	10 509	0.11%
$W$ (all shells)	0.090–0.380	0.67%
$T_{\text{rad}}$ (K)	6 861–12 291	0.28%
Escape fraction	~40%	< 1%

### 3.5 Bayesian inference

Given an observed spectrum  $\mathbf{F}_{\text{obs}}$  and the emulator prediction  $\mathbf{F}_{\text{emu}}(\boldsymbol{\theta})$ , the likelihood is:

$$\ln \mathcal{L}(\boldsymbol{\theta}) = -\frac{1}{2} \sum_i \frac{[F_{\text{obs},i} - F_{\text{emu},i}(\boldsymbol{\theta})]^2}{\sigma_i^2}, \quad (9)$$

where  $\sigma_i$  combines observational and emulator uncertainties with wavelength-dependent weighting:  $\sigma_{\text{UV}} = 2\sigma_{\text{base}}$ ,  $\sigma_{\text{NIR}} = 1.5\sigma_{\text{base}}$ , and  $\sigma_{\text{feat}} = 0.7\sigma_{\text{base}}$  in key feature windows, with  $\sigma_{\text{base}} = (\sigma_{\text{obs}}^2 + \sigma_{\text{emu}}^2)^{1/2}$ .

We employ three complementary inference methods:

- **MCMC** (Foreman-Mackey et al. 2013): 64 walkers, 2 000 burn-in steps, 5 000 production steps, yielding 320 000 posterior samples.
- **Nested sampling** (Speagle 2020): 500 live points,  $\Delta \ln Z = 0.1$  termination criterion, providing Bayesian evidence for model comparison.
- **Simulation-based inference** (SBI; Cranmer et al. 2020): Neural Posterior Estimation (NPE; Papamakarios et al. 2021) trained directly on simulation–observation pairs, yielding 10 000 posterior samples without explicit likelihood evaluation.

### 3.6 Posterior-based prior refinement (Stage 2.5)

The initial Stage 2 inference in 63 dimensions suffers from sparse coverage: 5 000 uniform samples in 63D provide only ~3.8 per cent overlap with the posterior. To address this, we narrow all 63 prior ranges to the [1, 99] percentile interval of the Stage 2 SBI posterior, expanded by a 30 per cent margin on each side and clamped to the original prior bounds. This achieves a total volume compression of  $\sim 10^{13.2}$ , ensuring that a new set of 5 000 training samples densely covers the region of interest.

## 4 APPLICATION TO SN 2011FE

We apply the full LUMINA/LUMINA-ML framework to SN 2011fe, the nearest and best-observed normal SN Ia in the modern era (Nugent et al. 2011). We use the spectrophotometric observation at phase  $-0.3$  d from *B*-maximum from the Nearby Supernova Factory (Pereira et al. 2013).

### 4.1 Validation against TARDIS

Table 1 summarizes the convergence of LUMINA relative to TARDIS for the SN 2011fe reference model. After 20 iterations with  $2 \times 10^5$  packets and identical input data, the maximum relative errors are 0.67 per cent in  $W$ , 0.28 per cent in  $T_{\text{rad}}$ , and 0.11 per cent in  $T_{\text{inner}}$ .

Fig. 1 compares the synthetic spectra from LUMINA (with varying NLTE start iterations) against TARDIS and the observed SN 2011fe spectrum. The deferred NLTE

**Table 2.** Computational performance of LUMINA ( $2 \times 10^5$  packets per iteration).

Configuration	Time/iter (s)	Speedup
CPU serial	7.5	1×
CPU + OpenMP (8 threads)	1.5	5×
GPU (RTX 5000 Ada)	0.72	10×
GPU + NLTE (cuBLAS)	+0.3 per NLTE iter	—
<i>Production run: <math>2 \times 10^6</math> pkts × 20 iters = 14 s (GPU)</i>		

**Table 3.** Stage 1 (15D) MCMC posterior for SN 2011fe at *B*-maximum. Uncertainties are 68 per cent credible intervals.

Parameter	Median	Unit
$\log L$	$42.80 \pm 0.01$	$\text{erg s}^{-1}$
$v_{\text{inner}}$	$7553 \pm 25$	$\text{km s}^{-1}$
$\log \rho_0$	$-13.99 \pm 0.01$	$\text{g cm}^{-3}$
$n_{\text{inner}}$	$-4.00 \pm 0.00$	—
$T_e/T_{\text{rad}}$	$0.84 \pm 0.01$	—
$v_{\text{core}}$	$9782 \pm 41$	$\text{km s}^{-1}$
$v_{\text{wall}}$	$15\,938 \pm 53$	$\text{km s}^{-1}$
$X_{\text{Fe,core}}$	$0.050 \pm 0.000$	—
$X_{\text{Si,wall}}$	$0.574 \pm 0.004$	—
$v_{\text{break}}$	$21\,982 \pm 20$	$\text{km s}^{-1}$
$n_{\text{outer}}$	$-5.64 \pm 0.14$	—
$t_{\text{exp}}$	$10.1 \pm 0.1$	d
$X_{\text{Fe,wall}}$	$0.059 \pm 0.002$	—
$X_{\text{Ni}}$	$0.237 \pm 0.004$	—
$X_{\text{Fe,outer}}$	$0.062 \pm 0.002$	—

strategy with  $N_{\text{start}} = 5$  provides the best overall agreement, reproducing the  $\text{Si II } \lambda 6355$  trough velocity to within  $74 \text{ km s}^{-1}$  of TARDIS and yielding an RMS residual of 0.163 relative to the observed

### 4.2 Performance benchmarks

Table 2 summarizes the computational performance of LUMINA across hardware configurations.

### 4.3 Stage 1: 15D global parameter inference

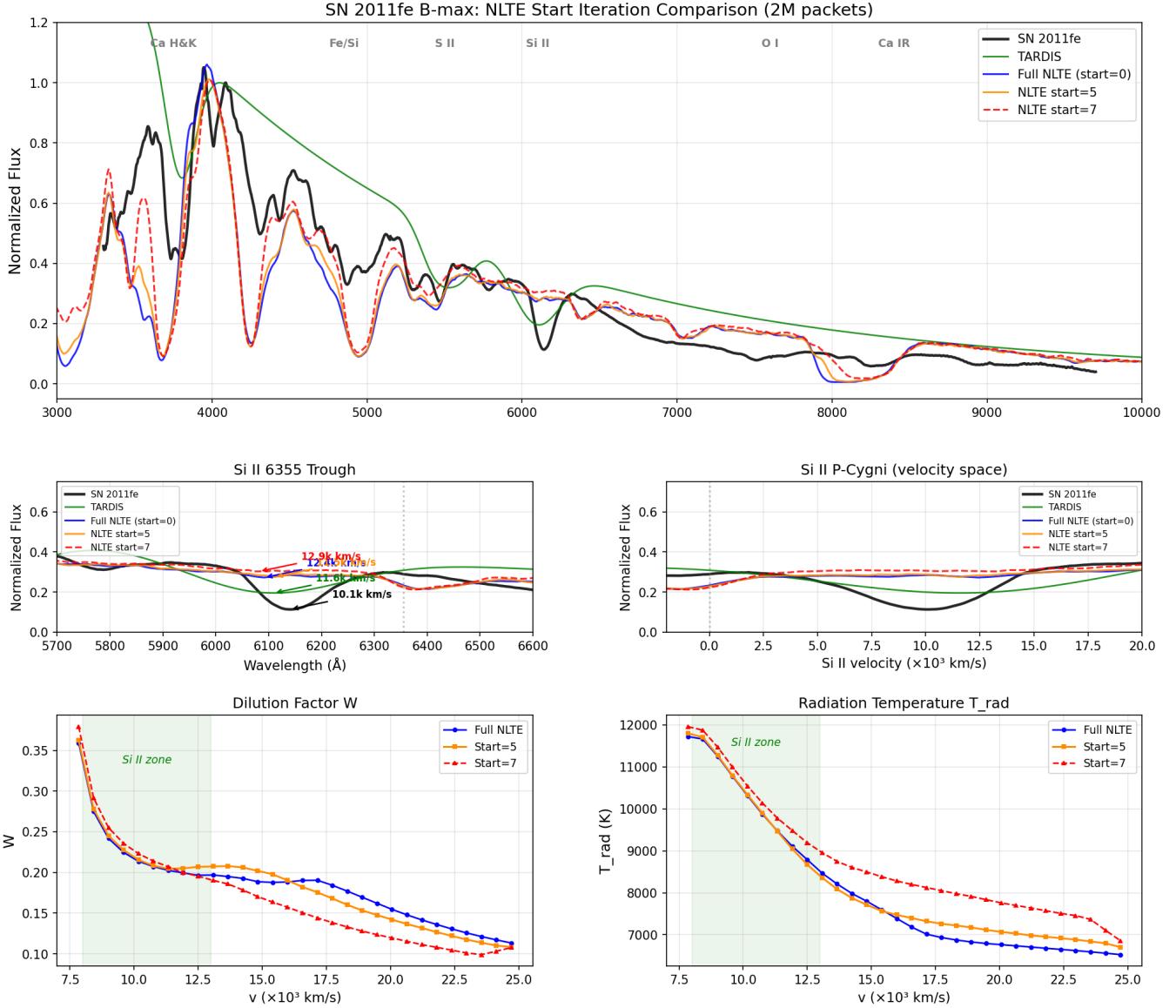
The Stage 1 emulator is trained on 25 000 LUMINA simulations (22 500 training, 2 500 validation) spanning the 15-dimensional prior. After 374 epochs, the validation loss converges with median spectral RMS of 0.36 per cent (Fig. 3).

Fig. 4 shows the 15D MCMC posterior for SN 2011fe. The posterior is well-constrained for most parameters, with tight correlations between luminosity, photospheric velocity, and density. Table 3 lists the median and 68 per cent credible intervals.

Fig. 5 shows the posterior predictive spectrum compared to the observed SN 2011fe spectrum. The emulator median faithfully reproduces the major spectral features, including  $\text{Si II } \lambda 6355$ ,  $\text{S II W}$ ,  $\text{Ca II H&K}$ , and the Fe/Co/Ni UV blanketing.

### 4.4 Method comparison

Fig. 6 compares the marginalized posteriors from MCMC, nested sampling, and SBI. The MCMC and dynesty posteriors are consistent and tightly constrained. The SBI posterior is systematically broader, as expected for an amortized method that does not assume a fixed



**Figure 1.** NLTE deferred start iteration comparison for SN 2011fe at  $B$ -maximum ( $2 \times 10^6$  packets). *Top:* Full spectrum with observed data (black), TARDIS reference (green), and LUMINA NLTE with start iterations 0 (blue), 5 (orange), and 7 (red dashed). *Middle left:* Zoom on the  $\text{Si II}$   $\lambda 6355$  trough showing velocity offsets. *Middle right:*  $\text{P-Cygni}$  profile in velocity space. *Bottom:* Dilution factor  $W$  and radiation temperature  $T_{\text{rad}}$  vs velocity, with the  $\text{Si II}$  formation zone shaded.

likelihood functional form, but the medians are generally consistent within the MCMC credible intervals.

#### 4.5 Stage 2: 63D abundance tomography

The Stage 2 analysis extends the parameter space to 63 dimensions, resolving the ejecta into six radial zones (shells 0–4, 5–9, ..., 25–29) with eight free elemental abundances per zone. The 15 physical parameters are relaxed by  $\pm 10$  per cent around the Stage 1 MCMC median.

The Stage 2 emulator is trained on 5 000 NLTE simulations (deferred start = 5) and achieves a median spectral RMS of 0.39 per cent. Fig. 7 shows the marginalized SBI posterior for the 15 physical parameters, and Fig. 8 shows the posterior predictive spectrum.

The Stage 2 SBI posterior medians for the physical parameters are

broadly consistent with Stage 1 (Table 4), with  $\log L = 42.79 \text{ erg s}^{-1}$  and  $v_{\text{inner}} = 7912 \text{ km s}^{-1}$ . The composition parameters reveal a stratified structure: the innermost zones (0–1) are iron-peak dominated ( $X_{\text{Fe}} \sim 0.2\text{--}0.7$ ,  $X_{\text{Ni}} \sim 0.14\text{--}0.40$ ), while the outer zones show increasing intermediate-mass element fractions.

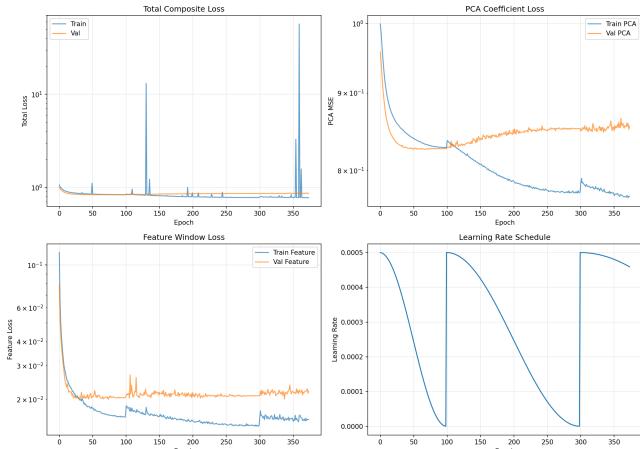
## 5 DISCUSSION

### 5.1 Physical interpretation

The inferred parameters for SN 2011fe are physically reasonable. The luminosity  $\log L = 42.80 \text{ erg s}^{-1}$  ( $L = 6.3 \times 10^{42} \text{ erg s}^{-1}$ ) is consistent with a normal SN Ia near  $B$ -maximum powered by  $\sim 0.6 M_{\odot}$  of  $^{56}\text{Ni}$ . The photospheric velocity  $v_{\text{inner}} \approx 7500\text{--}7900 \text{ km s}^{-1}$  is some-



**Figure 2.** Comparison of LUMINA NLTE spectra with TARDIS and observed SN 2011fe at *B*-maximum. *Top:* Normalized flux spectra with key features annotated. *Bottom:* Residuals relative to observation. Full NLTE (blue): RMS = 0.168; deferred NLTE (red): RMS = 0.144; TARDIS (green): RMS = 0.161.



**Figure 3.** Training history for the Stage 1 emulator: total loss, PCA MSE, feature loss, and learning rate schedule over 374 epochs.

what lower than the Si II-based estimate ( $\sim 10\,100\,\text{km s}^{-1}$ ), reflecting the fact that the Si II absorption forms above the photosphere in the optically thin regime.

**Table 4.** Comparison of physical parameter posteriors across stages and methods for SN 2011fe.

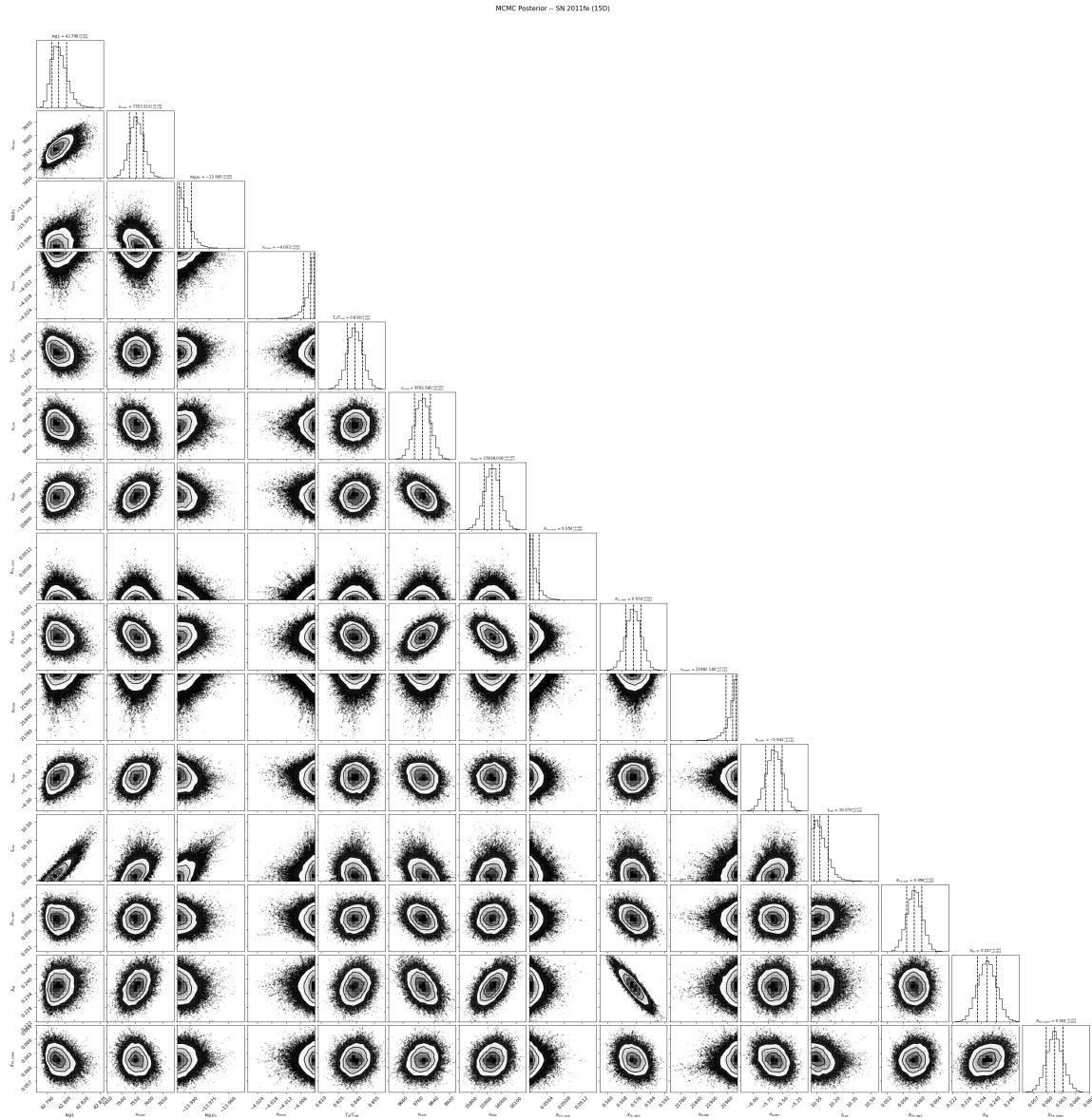
Parameter	S1 MCMC	S2 SBI	Unit
$\log L$	42.80	42.79	$\text{erg s}^{-1}$
$v_{\text{inner}}$	7553	7912	$\text{km s}^{-1}$
$\log \rho_0$	-13.99	-13.91	$\text{g cm}^{-3}$
$n_{\text{inner}}$	-4.00	-4.46	—
$T_e/T_{\text{rad}}$	0.84	0.83	—
$t_{\text{exp}}$	10.1	10.6	d
$X_{\text{Ni}}$	0.237	0.228	—

The nickel mass fraction  $X_{\text{Ni}} \sim 0.23$  (initial  $^{56}\text{Ni}$  at explosion) combined with the density profile implies a total  $^{56}\text{Ni}$  mass of  $\sim 0.5\text{--}0.7\,M_{\odot}$ , consistent with light-curve estimates for SN 2011fe.

## 5.2 Advantages of the multi-stage approach

The recursive refinement strategy offers several advantages:

- (i) *Computational efficiency:* Stage 1 (15D, 25k models) constrains global parameters; Stage 2 (63D, 5k models) refines compositions within the Stage 1 posterior; Stage 2.5 further concentrates training samples, achieving  $10^{13.2}\times$  volume compression.



**Figure 4.** MCMC corner plot for the Stage 1 (15D) posterior of SN 2011fe at  $B$ -maximum. Contours show 68 per cent and 95 per cent credible regions.

(ii) *Scalability*: The framework naturally extends to higher dimensions (Stage 3: 135D) without regenerating the full training set.

(iii) *Robustness*: Three independent inference methods (MCMC, nested sampling, SBI) provide cross-validation of posterior estimates.

### 5.3 Limitations

Several limitations merit discussion:

(i) The Sobolev approximation assumes that the velocity gradient dominates thermal broadening, which is well justified for SNe Ia at  $B$ -maximum ( $v/v_{\text{th}} \sim 1000$ ) but may break down in the innermost shells or at late times.

(ii) The one-dimensional geometry precludes modelling of polarization or asymmetric features.

(iii) The emulator accuracy (RMS  $\sim 0.4$  per cent) sets a floor

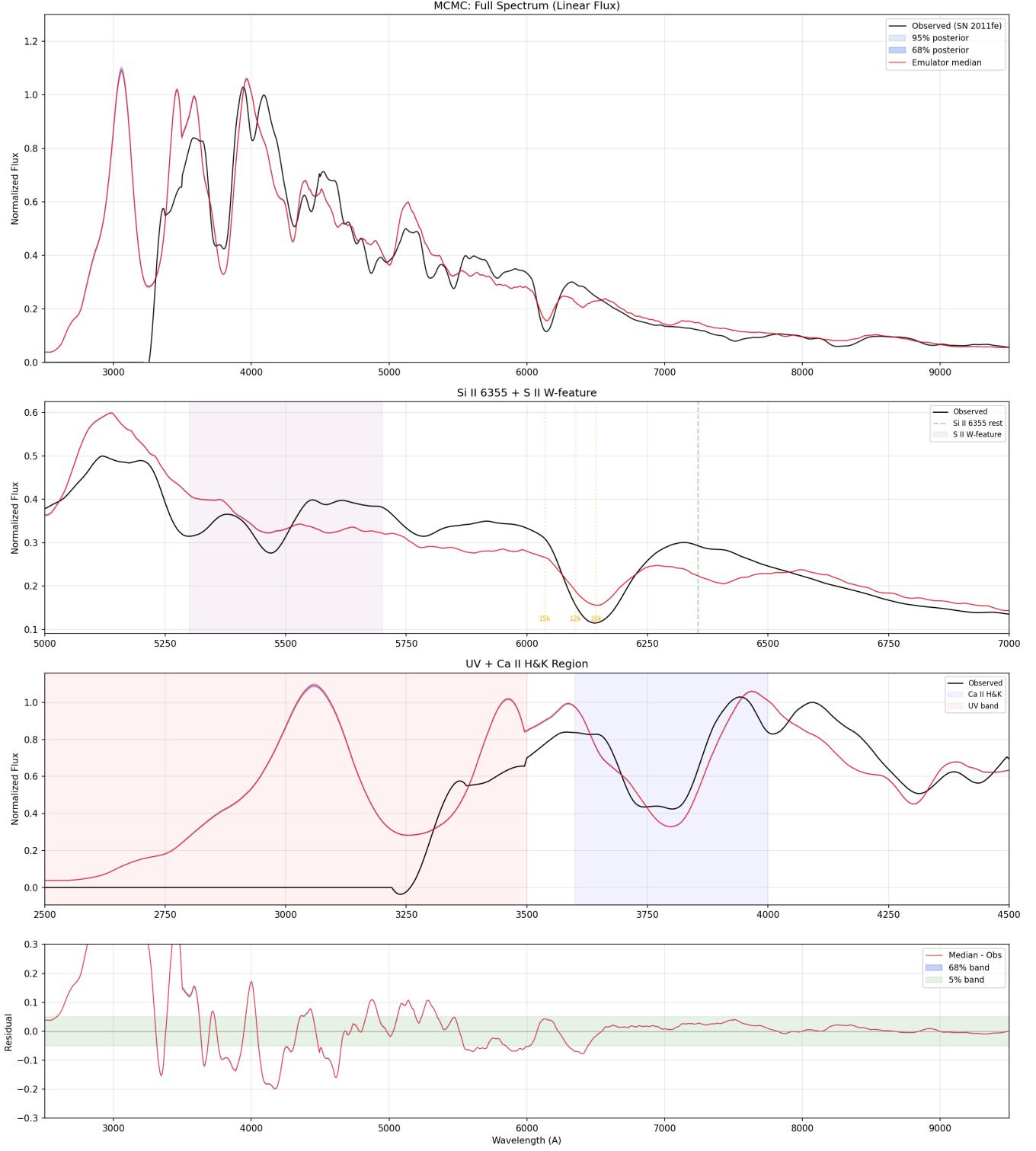
on the achievable posterior precision, though this is well below the observational uncertainties for most features.

(iv) The NLTE treatment covers only the dominant ions; rarer species ( $\text{C II}$ ,  $\text{Mg II}$ ) remain in LTE.

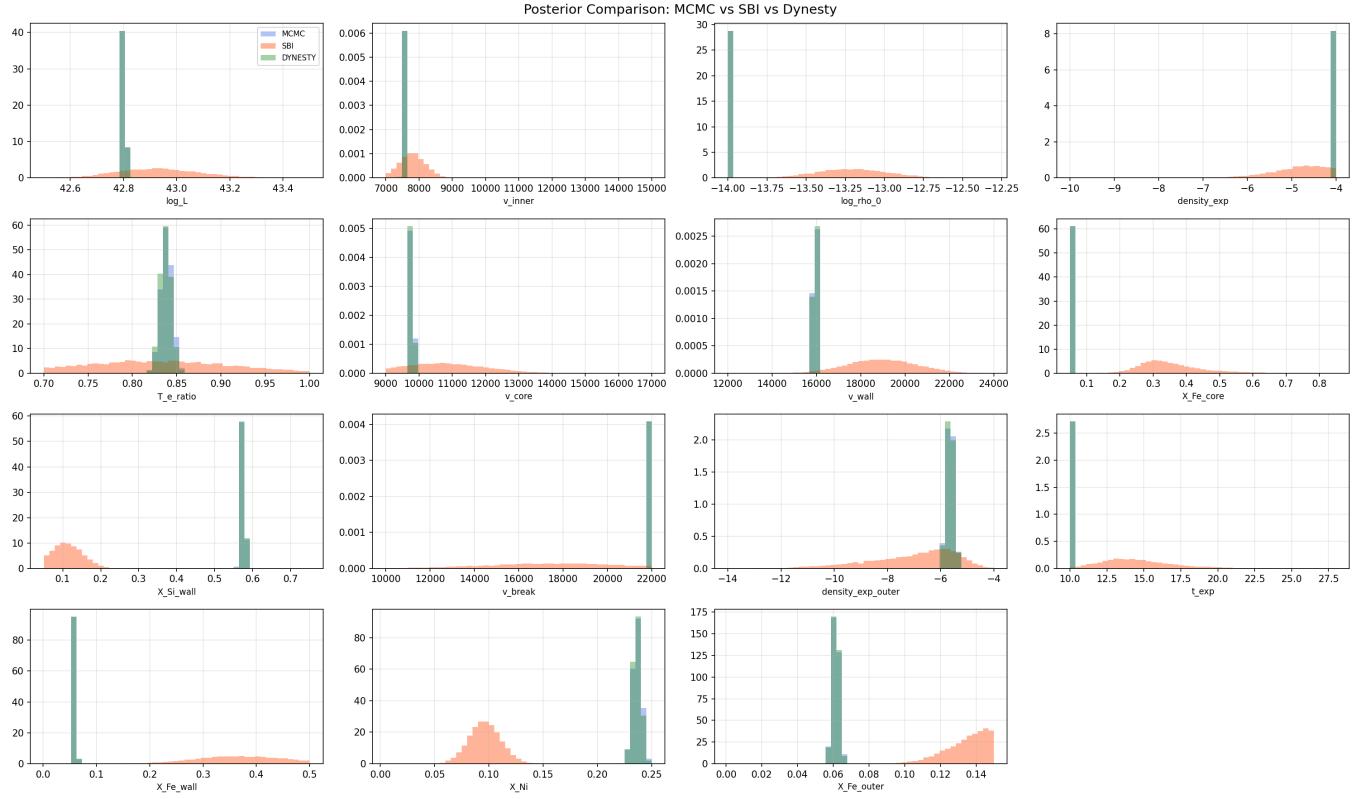
(v) The dilute Planck approximation  $J_{\nu} \approx W B_{\nu}(T_{\text{rad}})$  used in the NLTE solver may introduce systematic biases relative to a per-line  $J_{\nu}$  estimator.

### 5.4 Future work

Planned extensions include: (i) inclusion of  $\text{C II}$  and  $\text{Mg II}$  in the NLTE solver; (ii) a per-line  $J_{\nu}$  estimator () for improved macro-atom source functions; (iii) multi-epoch spectral fitting to leverage time-series constraints; (iv) application to a larger sample of SNe Ia for population-level abundance studies; and (v) Stage 3 refinement to 135 dimensions with 15 radial zones.



**Figure 5.** Stage 1 MCMC posterior predictive spectrum for SN 2011fe. The observed spectrum (black) is compared with the emulator median (blue) and 68/95 per cent posterior bands (shaded). Individual panels show the full spectrum and zooms on key features.



**Figure 6.** Comparison of Stage 1 marginalized posteriors from three inference methods: MCMC (blue), nested sampling (green), and SBI (orange) for SN 2011fe at  $B$ -maximum.

## 6 CONCLUSIONS

We have presented LUMINA, a GPU-accelerated Monte Carlo radiative transfer code for Type Ia supernova spectral synthesis, and LUMINA-ML, a multi-stage machine learning pipeline for Bayesian parameter inference. The key results are:

- (i) LUMINA reproduces the TARDIS reference solution to within 0.11 per cent ( $T_{\text{inner}}$ ), 0.67 per cent ( $W$ ), and 0.28 per cent ( $T_{\text{rad}}$ ), while achieving a 10 $\times$  GPU speedup.
- (ii) The NLTE solver for six ion pairs (3 547 levels) with charge exchange coupling and deferred activation ( $N_{\text{start}} = 5$ ) matches the TARDIS Si II  $\lambda 6355$  trough velocity to within 74 km s $^{-1}$ .
- (iii) The LUMINA-ML neural network emulator achieves 0.36–0.39 per cent median spectral RMS, enabling posterior estimation over 15–63 parameters via MCMC, nested sampling, and SBI.
- (iv) Applied to SN 2011fe at  $B$ -maximum, the framework yields  $\log L = 42.80 \pm 0.01$  erg s $^{-1}$ ,  $v_{\text{inner}} = 7553 \pm 25$  km s $^{-1}$ ,  $t_{\text{exp}} = 10.1 \pm 0.1$  d, and spatially resolved abundance profiles consistent with the standard layered detonation picture.
- (v) The posterior-based prior refinement (Stage 2.5) achieves 10 $^{13.2}\times$  volume compression, enabling efficient 63D inference with only 5 000 training models.

The full pipeline — from raw simulation to posterior parameter estimates — runs in hours on a single GPU, opening the door to systematic spectroscopic analysis of large SN Ia samples.

## ACKNOWLEDGEMENTS

We thank the TARDIS development team for their open-source radiative transfer framework, which served as the physics reference for LUMINA. Atomic line data are from the Kurucz CD-ROM No. 23 (Kurucz 1995) and the CHIANTI database. The SN 2011fe spectrophotometry is from the Nearby Supernova Factory (Pereira et al. 2013). Computations were performed on NVIDIA RTX 5000 Ada GPUs.

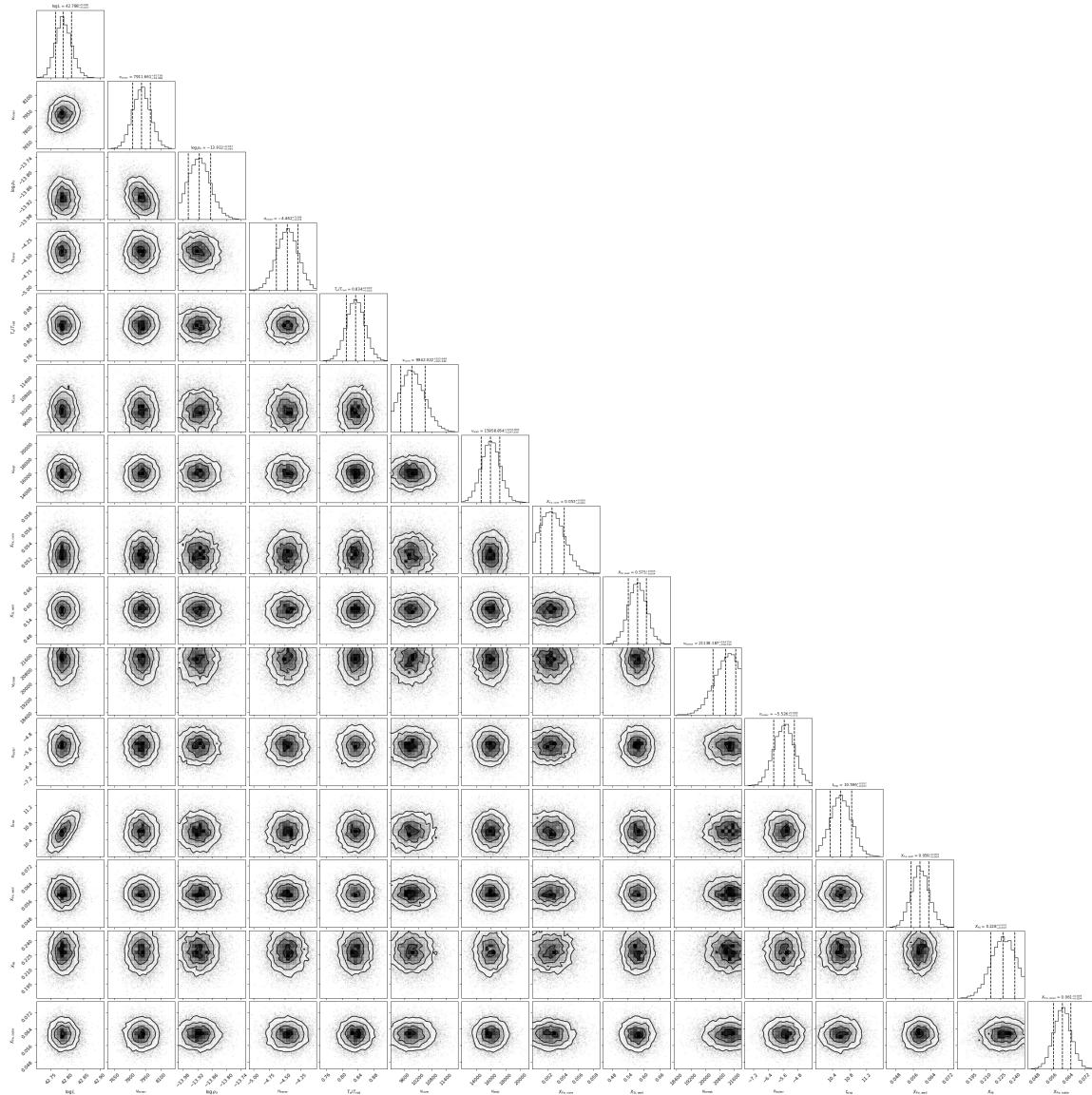
## DATA AVAILABILITY

The LUMINA source code and LUMINA-ML pipeline are available at . The trained emulator weights, posterior samples, and preprocessed data will be shared upon reasonable request.

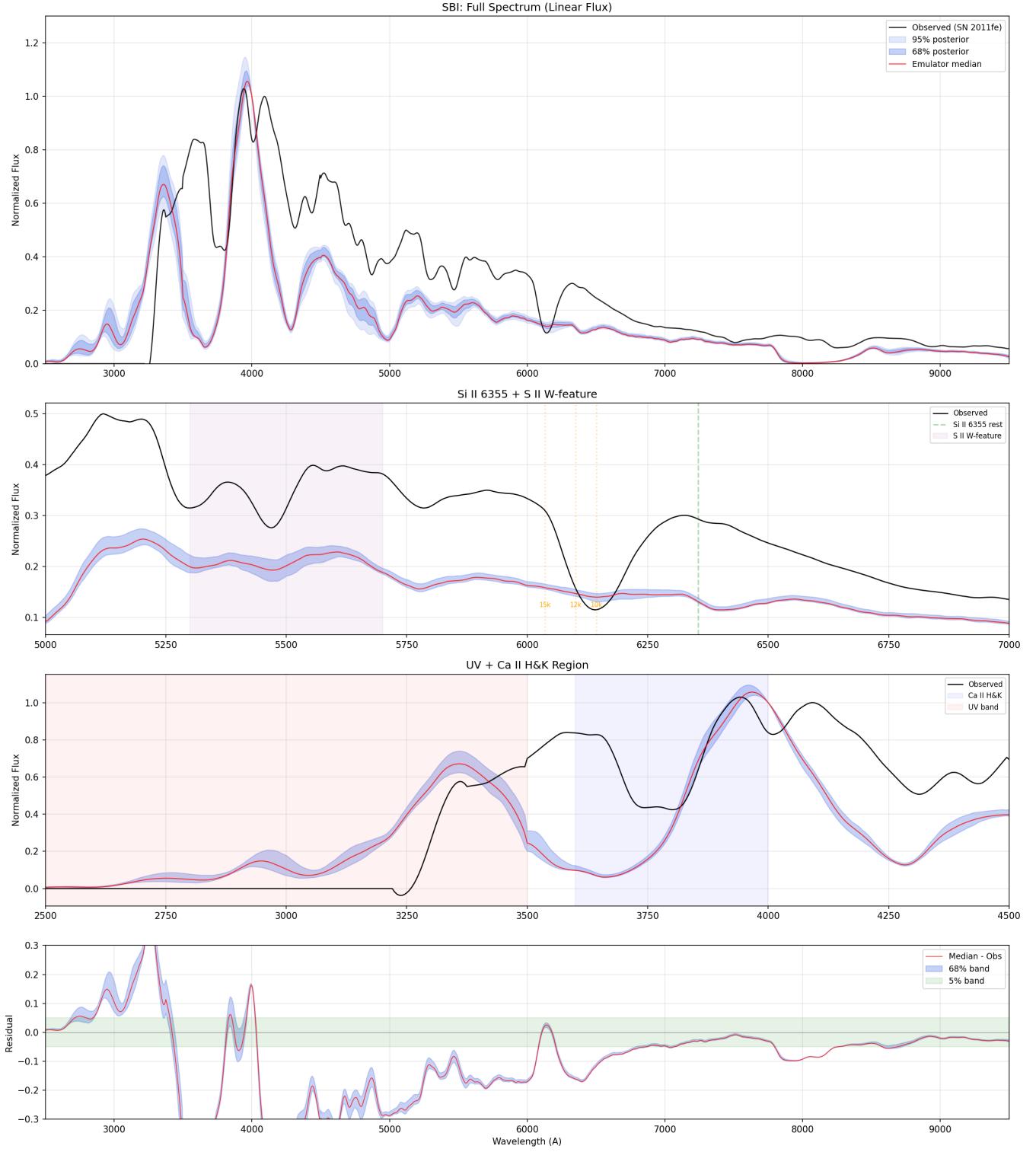
## REFERENCES

- Blackman D., Vigna S., 2021, *ACM Transactions on Mathematical Software*, 47
- Branch D., Dang L. C., Hall N., et al., 2006, *PASP*, 118, 560
- Cranmer K., Brehmer J., Louppe G., 2020, *Proceedings of the National Academy of Sciences*, 117, 30055
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
- Kasen D., Thomas R. C., Nugent P., 2006, *ApJ*, 651, 366
- Kerzendorf W. E., Sim S. A., 2014, *MNRAS*, 440, 387
- Kerzendorf W. E., Vogl C., Buchner J., et al., 2021, arXiv e-prints
- Kurucz R. L., 1995, Kurucz CD-ROM No. 23

SBI Posterior – SN 2011fe (63D, showing 15 physical)

**Figure 7.** Stage 2 SBI corner plot for the 15 physical parameters of SN 2011fe, marginalized over 48 composition parameters.

- Lucy L. B., 1999, *A&A*, 345, 211  
 Lucy L. B., 2002, *A&A*, 384, 725  
 Lucy L. B., 2003, *A&A*, 403, 261  
 Nomoto K., Thielemann F.-K., Yokoi K., 1984, *ApJ*, 286, 644  
 Nugent P. E., Sullivan M., Cenko S. B., et al., 2011, *Nature*, 480, 344  
 O'Brien J. T., Stevance H. F., Sim S. A., 2021, *MNRAS*, 503, 4942  
 Papamakarios G., Nalisnick E., Rezende D. J., Mohamed S., Lakshmi-  
     narayanan B., 2021, *Journal of Machine Learning Research*, 22, 1  
 Pereira R., Thomas R. C., Aldering G., et al., 2013, *A&A*, 554, A27  
 Perlmutter S., Aldering G., Goldhaber G., et al., 1999, *ApJ*, 517, 565  
 Phillips M. M., 1993, *ApJ*, 413, L105  
 Riess A. G., Filippenko A. V., Challis P., et al., 1998, *AJ*, 116, 1009  
 Sim S. A., 2007, *MNRAS*, 375, 154  
 Sobolev V. V., 1960, Cambridge: Harvard University Press  
 Speagle J. S., 2020, *MNRAS*, 493, 3132  
 Vogl C., Sim S. A., Noebauer U. M., Kerzendorf W. E., Hillebrandt W., 2020,  
     *A&A*, 633, A88  
 van Regemorter H., 1962, *ApJ*, 136, 906



**Figure 8.** Stage 2 SBI posterior predictive spectrum for SN 2011fe, with 63-dimensional composition resolution.