
Speeding up astrochemical reaction networks with autoencoders and neural ODEs

Immanuel Sulzer

Interdisciplinary Center
for Scientific Computing,
University of Heidelberg,
Im Neuenheimer Feld 205,
D-69120 Heidelberg

`immanuel.sulzer@stud.uni-heidelberg.de`

Tobias Buck

Interdisciplinary Center
for Scientific Computing,
University of Heidelberg,
Im Neuenheimer Feld 205,
D-69120 Heidelberg

`tobias.buck@iwr.uni-heidelberg.de`

Abstract

In astrophysics, solving complex chemical reaction networks is essential but computationally demanding due to the high dimensionality and stiffness of the ODE systems. Traditional approaches for reducing computational load are often specialized to specific chemical networks and require expert knowledge. This paper introduces a machine learning-based solution employing autoencoders for dimensionality reduction and a latent space neural ODE solver to accelerate astrochemical reaction network computations. Additionally, we propose a cost-effective latent space linear function solver as an alternative to neural ODEs. These methods are assessed on a dataset comprising 29 chemical species and 224 reactions. Our findings demonstrate that the neural ODE achieves a 55x speedup over the baseline model while maintaining significantly higher accuracy by up to two orders of magnitude reduction in relative error. Furthermore, the linear latent model enhances accuracy and achieves a speedup of up to 4000x compared to standard methods.

1 Motivation

Chemistry plays a pivotal role in astrophysical simulations, influencing phenomena in the interstellar medium [11] and star formation processes [6]. The chemical composition of species not only impacts physical processes but also affects observational outcomes, particularly line emissions [e.g. 9]. However, solving chemical reaction networks poses challenges, given the need to handle numerous tightly coupled stiff ODEs, arising from distinct reaction timescales. Existing simplification techniques, such as species omission and grouping [16], instantaneous equilibrium assumption [7], or exploiting sparse reaction matrices [8], often rely heavily on specific chemical networks and may not always yield effective results. As an alternative, Grassi et al. demonstrated the effectiveness of autoencoders for dimensionality reduction in astrochemistry. Furthermore, the advent of neural ODEs [4, 12] has opened new possibilities for novel ODE solver methods in astrochemical modelling. Our approach seeks to integrate these techniques, aiming to create an easily implementable and efficient solver for chemical reaction networks. Our code is publicly available on Github: (URL: <https://github.com/Immi000/ChemicalNeuralODE>). The code is documented and can easily be adapted for reuse.

2 Latent space neural ODE solver

A neural ODE essentially uses a neural network (NN) f to model the derivative of a vector valued function $\mathbf{x}(t)$ that depends on a parameter t (time in our application): $\frac{d\mathbf{x}(t)}{dt} = f(\mathbf{x}(t), t)$. Our model for solving chemical reaction networks comprises three components: an encoder φ , the

neural ODE f , and a decoder ψ . The dynamics of the chemical system are characterized by a function $x(t)$, where only the initial state x_0 is known. The encoder transforms the initial state x_0 into the latent space initial state $z_0 = \varphi(x_0)$, which serves as the initial condition for the ODE: $\frac{dz(t)}{dt} = f(z(t))$ with $z(t=0) = z_0$ where $z(t)$ represents the latent space time evolution of the chemical system. $x(t)$ can therefore be calculated as:

$$x(t) = \psi \left(\varphi(x_0) + \int_{[t_0, t_1]} f(z(t)) dt \right) \quad (1)$$

where the integral can be evaluated using any ODE solver. For reasons of performance we have opted for the `odeint_adjoint`-method implemented by Chen with a Dormand-Prince solver [5], as it is capable of backpropagation via the adjoint sensitivity method [3, 4]. During experimentation, we noticed that some of the trajectories in the latent space are approximately linear (see figure 3). This means the latent space ODE is approximately given by a constant $\frac{dz(t)}{dt} = f(z(t)) \approx c$. This simplifies integration, obviating the need for a numerical solver or neural ODE. The solution becomes: $z(t) = \int c dt = ct + z_0$ enabling direct evaluation at all time points, significantly accelerating overall computation. We evaluate this approach as a "linear latent" model and compare it with the neural ODE approach. As a broader strategy, one can select any analytically integrable function (e.g., a polynomial) to approximate the latent space evolution, preserving the computational advantages of the linear model while enabling higher complexity in the latent space. As the model's ultimate training objective is to discover a latent space representation with greatly simplified dynamics, it draws inspiration from manifold learning and sparse identification of nonlinear dynamics [2].

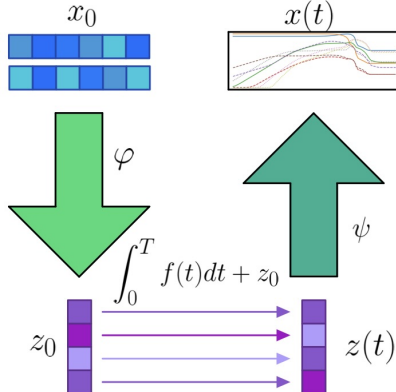


Figure 1: The encoder φ encodes the initial state x_0 into its latent space equivalent z_0 . The time evolution $z(t)$ is then calculated by integrating the neural ODE f with the initial condition z_0 . The calculated latent space evolution of $z(t)$ is then transformed back into the real space time evolution of abundances $x(t)$ via the decoder ψ .

2.1 Training data

Our training data is based on the chemical network `osu_09_2008` and has been generated following [10]. This network consists of 224 reactions between 29 species and assumes other physical parameters such as temperature ($T = 50$ K), cosmic ray ionization rate ($\zeta = 10^{-16} s^{-1}$) and total density ($n_{tot} = 10^4 cm^{-3}$) to be constant. We simulate the species evolution 500 times over 100 time steps, varying the initial conditions¹. The chemical network solves the reactions within a logarithmic space, which is min-max normalized to the range $[-1, 1]$. The initial state is assumed to be predominantly molecular, with n_{H_2} equal to n_{tot} . n_C and n_{C+} are randomized in log space, ranging between $10^{-6}n_{tot}$ and $10^{-3}n_{tot}$. Molecular Oxygen is initialized to equal $n_O = n_C + n_{C+}$. The initialization of electrons ensures that the system maintains charge neutrality, while all other species commence at an initial value of $10^{-20}n_{tot}$. The time evaluation points span an interval of 10^8 years and are evenly distributed on a logarithmic scale, subsequently min-max normalized to $[0, 1]$.

2.2 Architecture design and loss function

Our final model consists of three NNs: the encoder φ , the decoder ψ , and the neural ODE f . The **encoder** φ is a fully connected feedforward NN with four hidden layers. The input layer has 29 dimensions, one for each of the 29 chemical species. The four hidden layers have dimensions of 64, 32, 16, and 8, respectively, and the output layer has five dimensions, corresponding to the latent space dimensionality for this experiment. The activation function used in both the encoder and decoder is $\tanh$. The **decoder** ψ mirrors the architecture of the encoder in reverse order. The **neural ODE** f consists of a fully connected feedforward NN with an input layer, 5 hidden layers of width 64, and an output layer. The activation function after every layer (except the last) is $\tanh$. To allow for a wider range of output gradients, a scaled version of $\tanh$ is used, including a learnable scalar parameter s : $\frac{dz(t)}{dt} = s \cdot \tanh(\mathbf{NN}(z)/s)$. This modification extends the output of the NN to the range of $[-s, s]$, offering greater flexibility in gradient values and introducing a regularizing effect [19]. For model

¹The code for generating the data can be found in the repository of [10]

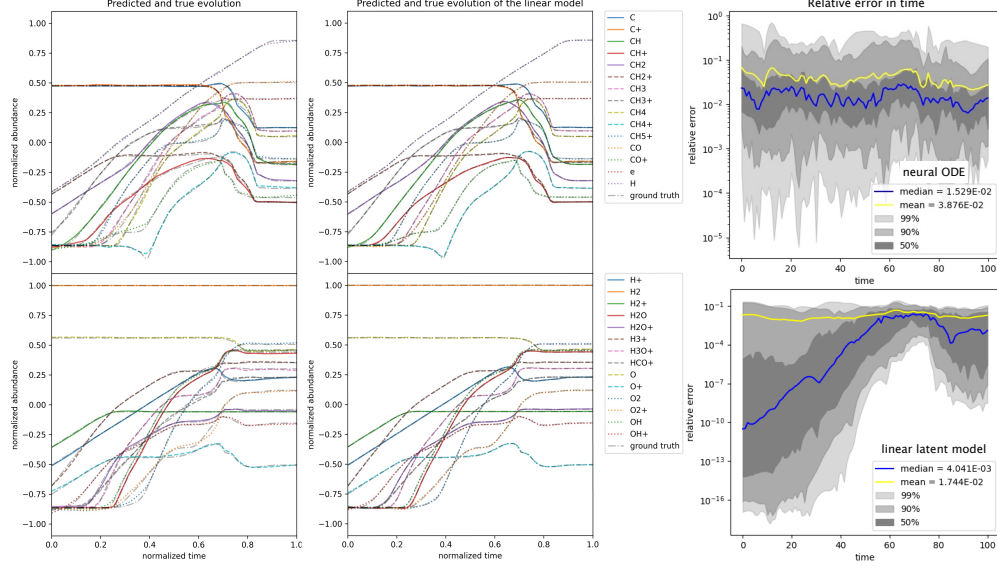


Figure 2: The ground truth (grey) and the predicted abundances (color) from the neural ODE model (left) and the linear latent space model (middle). The right panels show the relative error over time for the neural ODE model (upper panel) and the linear latent model (lower panel).

training, five loss functions are defined: The **reconstruction loss** (L_0) measures the discrepancy between the predicted trajectory $\mathbf{x}'(t)$ and the ground truth $\mathbf{x}(t)$, computed as $L_0 = \|\mathbf{x}(t) - \mathbf{x}'(t)\|^2$. The **identity loss** (L_1) ensures that the combination of the encoder φ and decoder ψ approximates the identity. It is calculated as $L_1 = \|\mathbf{x}(t) - \psi(\varphi(\mathbf{x}(t)))\|^2$. The **mass conservation loss** (L_2), incorporating the species masses vector $\mathbf{m}$, ensures that the mass of the predicted trajectory aligns with the mass of the ground truth at each time point. It is computed as $L_2 = \|\mathbf{m} \cdot \mathbf{x}(t) - \mathbf{m} \cdot \mathbf{x}'(t)\|^2$. **Gradient losses** (L_3 and L_4) enforce the predicted trajectories to closely follow the ground truth by comparing first and second-order gradients in real space. L_3 measures the discrepancy in first-order gradients, and L_4 examines second-order gradients. They are defined as $L_3 = \|\frac{d\mathbf{x}(t)}{dt} - \frac{d\mathbf{x}'(t)}{dt}\|^2$ and $L_4 = \|\frac{d^2\mathbf{x}(t)}{dt^2} - \frac{d^2\mathbf{x}'(t)}{dt^2}\|^2$. After a burn-in phase of ten epochs, the losses are normalized to unity and the L_0 loss is weighted by 100. Optimization is performed using the Adam optimizer [13] with a cosine-annealing learning rate scheduler, varying the learning rate between 10^{-3} and 10^{-5} .

3 Results

The models are evaluated on a test set of 50 unseen trajectories. Fig.2 shows one sample from the test set alongside the predictions from both models: the neural ODE (left panels) and the linear latent model (middle panels). The predicted trajectories closely match the ground truth trajectories. For certain species such as H, H_2O , or H_3^+ , there is hardly any discernible deviation while for other species like C or CH_4 , some minor discrepancies from the ground truth are noticeable. Notably, regions with moderate to strong curvature produce larger errors than flatter trajectories. Quantitatively, the deviations correspond to a Root Mean Square Error (RMSE) of $9 \cdot 10^{-3}$ for the neural ODE model and $3 \cdot 10^{-3}$ for the linear latent model. Additionally, the relative error (rightmost panels in Fig. 2) for the neural ODE trajectory ranges from $5.96 \cdot 10^{-8}$ to 0.58, with a median relative error of 0.015. In contrast, the linear latent model exhibits relative errors ranging between 10^{-16} and ~ 0.1 , with a median relative error of 0.004. In comparison, Grassi et al. employed a similar dimensionality reduction approach utilizing autoencoders but modelling reactions in latent space by explicitly defining a sparse reaction matrix. Evaluated on the same dataset, Grassi et al. achieve a logarithmic relative error $\langle \log(\Delta x) \rangle$ ranging between -2.96 and 0.44 . In contrast, our neural ODE model achieves a log relative error ranging from -4.41 to -1.03 , while our linear latent space model reaches a log relative error ranging from -5.09 to -1.64 . Overall, we have achieved significantly improved reconstruction results, with relative errors several orders of magnitude lower than those reported by [10]. The linear latent space model, in particular, achieves a relative error

Table 1: Performance measurements of different solver methods.

Model	integration time	faster than <code>solve_ivp</code>	faster than Neural ODE
<code>scipy solve_ivp</code>	1930 ms	-	0.018
Grassi et al.	29.7 ms	65	1.18
Neural ODE	34.9 ms	55	-
Linear Model	0.45 ms	4270	77

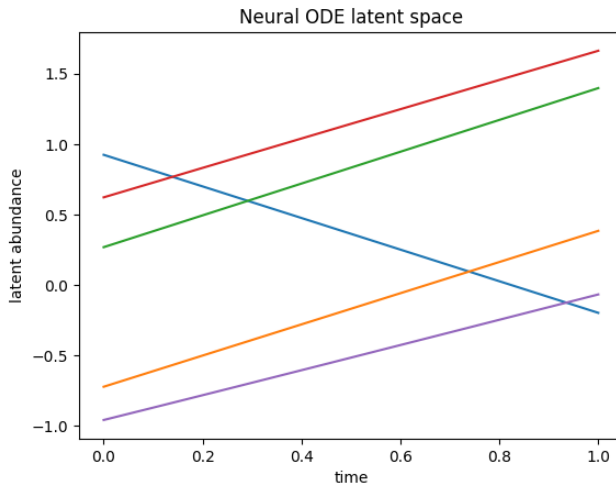


Figure 3: The predicted evolution in the latent space as calculated by the neural ODE (left) and the linear latent model (right). Even though the latent space trajectories of the neural ODE on the left were not constrained to be straight, in some cases the optimal solution found by the optimizer resulted in the simple trajectories shown. This phenomenon inspired the simpler and more performant linear latent model.

roughly one order of magnitude lower than that of the neural ODE model. This is further reflected in the logarithmic mean average error across all species, where Grassi et al. achieved -0.41 , while our neural ODE model achieves -2.40 , and the linear model reports -2.94 . This represents an improvement of approximately two orders of magnitude compared to previous results.

3.1 Speed

Performance measurements of various solver methods were conducted, measuring the time required for a model to predict the evolution of an initial abundance from the test set over 100 time steps. The standard `solve_ivp` integration, which was used with a multi-step variable-order BDF solver and employed in generating the training data, was executed on a Dual Intel Xenon Gold 6254 processor. In contrast, all other methods were run on an Nvidia RTX 2080ti GPU. In comparison to the reported performance by Grassi et al., who achieved a speed-up factor of 65 over `solve_ivp`, our neural ODE model achieves a speed-up factor of 55. Remarkably, the linear latent space model demonstrates a substantial speed-up, reaching a factor of 4270 significantly outperforming both the neural ODE model and the sparse solver by a factor of approximately 77.

4 Limitations & Outlook

The impressive results of the model should be considered in light of several limitations: The model’s performance is constrained by a limited range of initial conditions. The absence of variation in temperature, cosmic ray ionization rate, and total density as initial conditions means that the trajectories cover only a limited portion of the phase space. In reality, these parameters change over time, and incorporating such changes into the chemistry solver is challenging, as suggested by related

work [1]. The chemical reaction network considered in this study includes 224 reactions and 29 species. However, real-world astrochemical systems can involve much larger networks with over 400 species [e.g. 14]. It remains uncertain whether the methods tested in this study would perform effectively on such extensive networks. While the overall accuracy is high, there are regions and species, particularly those with low abundances and high reaction rates, where accurate predictions may not be achieved. The dimensionality reduction approach relies on data being projectable to a lower-dimensional manifold while preserving information about reactions. Questions remain about the optimal choice of latent space dimensions and the identification of the latent dynamics.

Looking forward, several promising directions for further research and exploration can be considered: Investigating the mathematical properties of the autoencoder, particularly why latent space trajectories tend to become straight lines without explicit constraints, could provide valuable insights into the model’s behavior. Further exploration is needed to understand the observed latent space trend where, in five dimensions, two lines typically have negative slopes while the others have positive slopes, albeit with similar magnitudes. Exploring advancements in machine learning for modeling physical systems and solving linear equations can enhance the model’s capabilities. Approaches such as the LM-architecture proposed by [15], coordinate-based neural operators for solving PDEs [18], and sparse identification methods [2] are promising areas for investigation. Investigating alternative methods, such as continuous-time echo state networks for modeling disturbed systems [17], can expand the toolbox for solving complex chemical simulations. Exploring these methods may provide complementary or alternative solutions to the challenges posed by large chemical networks.

In summary, while the current model has shown impressive results, there are still opportunities for improvement and avenues for further research to address the limitations and enhance its capabilities in solving complex astrochemical simulations.

Broader impact statement

This work purely aims to aid scientific algorithm design to speed-up ODE calculations. Therefore the authors are not aware of any immediate ethical or societal implications of this work. Thinking more broadly, many relevant systems in a broad range of science can be described by coupled ODEs and as such our approach of accelerating their solution with the help of ML might have beneficial impact beyond astrophysical modelling. This may contribute to a broader application and development of such algorithms and in the long run might help to save energy when modelling complex systems.

References

- [1] Lorenzo Branca and Andrea Pallottini. Neural networks: solving the chemistry of the interstellar medium. *Monthly Notices of the Royal Astronomical Society*, 518(4):5718–5733, December 2022. ISSN 0035-8711, 1365-2966. doi: 10.1093/mnras/stac3512. URL <http://arxiv.org/abs/2211.15688>. arXiv:2211.15688 [astro-ph].
- [2] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, April 2016. doi: 10.1073/pnas.1517384113. URL <https://www.pnas.org/doi/10.1073/pnas.1517384113>. Publisher: Proceedings of the National Academy of Sciences.
- [3] Ricky T. Q. Chen. torchdiffeq, 2018. URL <https://github.com/rtqichen/torchdiffeq>.
- [4] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural Ordinary Differential Equations, December 2019. URL <http://arxiv.org/abs/1806.07366>. arXiv:1806.07366 [cs, stat].
- [5] J.R. Dormand and P.J. Prince. A family of embedded Runge-Kutta formulae. *Journal of Computational and Applied Mathematics*, 6(1):19–26, March 1980. ISSN 03770427. doi: 10.1016/0771-050X(80)90013-3. URL <https://linkinghub.elsevier.com/retrieve/pii/0771050X80900133>.
- [6] S. C. O. Glover and A.-K. Jappsen. Star formation at very low metallicity. I: Chemistry and cooling at low densities. *The Astrophysical Journal*, 666(1):1–19, September 2007. ISSN

- 0004-637X, 1538-4357. doi: 10.1086/519445. URL <http://arxiv.org/abs/0705.0182>. arXiv:0705.0182 [astro-ph].
- [7] S. C. O. Glover, C. Federrath, M.-M. Mac Low, and R. S. Klessen. Modelling CO formation in the turbulent interstellar medium. *Monthly Notices of the Royal Astronomical Society*, 404(1):2–29, May 2010. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2009.15718.x. URL <https://doi.org/10.1111/j.1365-2966.2009.15718.x>.
 - [8] T. Grassi, S. Bovino, D. Schleicher, and F. A. Gianturco. Chemical complexity in astrophysical simulations: optimization and reduction techniques. *Monthly Notices of the Royal Astronomical Society*, 431(2):1659–1668, May 2013. ISSN 0035-8711. doi: 10.1093/mnras/stt284. URL <https://doi.org/10.1093/mnras/stt284>.
 - [9] T. Grassi, S. Bovino, D. R. G. Schleicher, J. Prieto, D. Seifried, E. Simoncini, and F. A. Gianturco. KROME - a package to embed chemistry in astrophysical simulations. *Monthly Notices of the Royal Astronomical Society*, 439(3):2386–2419, March 2014. ISSN 0035-8711, 1365-2966. doi: 10.1093/mnras/stu114. URL <http://arxiv.org/abs/1311.1070>. arXiv:1311.1070 [astro-ph].
 - [10] T. Grassi, F. Nauman, J. P. Ramsey, S. Bovino, G. Picogna, and B. Ercolano. Reducing the complexity of chemical networks via interpretable autoencoders. *Astronomy & Astrophysics*, 668:A139, December 2022. ISSN 0004-6361, 1432-0746. doi: 10.1051/0004-6361/202039956. URL <http://arxiv.org/abs/2104.09516>. arXiv:2104.09516 [astro-ph, physics:physics].
 - [11] D. Hollenbach and C. F. McKee. Molecule formation and infrared emission in fast interstellar shocks. I. Physical processes. *The Astrophysical Journal Supplement Series*, 41:555–592, November 1979. ISSN 0067-0049. doi: 10.1086/190631. URL <https://ui.adsabs.harvard.edu/abs/1979ApJS...41..555H>. ADS Bibcode: 1979ApJS...41..555H.
 - [12] Patrick Kidger. On Neural Differential Equations, February 2022. URL <http://arxiv.org/abs/2202.02435>. arXiv:2202.02435 [cs, math, stat].
 - [13] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization, January 2017. URL <http://arxiv.org/abs/1412.6980>. arXiv:1412.6980 [cs].
 - [14] Y. H. Le Teuff, T. J. Millar, and A. J. Markwick. The UMIST database for astrochemistry 1999. *Astronomy and Astrophysics Supplement Series*, 146(1):157–168, October 2000. ISSN 0365-0138, 1286-4846. doi: 10.1051/aas:2000265. URL <http://aas.aanda.org/10.1051/aas:2000265>.
 - [15] Yiping Lu, Aoxiao Zhong, Quanzheng Li, and Bin Dong. Beyond Finite Layer Neural Networks: Bridging Deep Architectures and Numerical Differential Equations, March 2020. URL <http://arxiv.org/abs/1710.10121>. arXiv:1710.10121 [cs, stat].
 - [16] Richard P. Nelson and William D. Langer. On the Stability and Evolution of Isolated Bok Globules. *The Astrophysical Journal*, 524(2):923, October 1999. ISSN 0004-637X. doi: 10.1086/307823. URL <https://iopscience.iop.org/article/10.1086/307823/meta>. Publisher: IOP Publishing.
 - [17] Ciaran Roberts, José Daniel Lara, Rodrigo Henriquez-Auba, Matthew Bossart, Ranjan Anantharaman, Chris Rackauckas, Bri-Mathias Hodge, and Duncan S. Callaway. Continuous-time echo state networks for predicting power system dynamics. *Electric Power Systems Research*, 212:108562, November 2022. ISSN 0378-7796. doi: 10.1016/j.epsr.2022.108562. URL <https://www.sciencedirect.com/science/article/pii/S0378779622006587>.
 - [18] Louis Serrano, Lise Le Boudec, Armand Kassaï Koupaï, Thomas X. Wang, Yuan Yin, Jean-Noël Vittaut, and Patrick Gallinari. Operator Learning with Neural Fields: Tackling PDEs on General Geometries, June 2023. URL <http://arxiv.org/abs/2306.07266>. arXiv:2306.07266 [cs].
 - [19] Kwok Sun Tang and Matthew Turk. Reduced Order Model for Chemical Kinetics: A case study with Primordial Chemical Network, July 2022. URL <http://arxiv.org/abs/2207.07159>. arXiv:2207.07159 [astro-ph, physics:physics].