Emulating Cosmic Structure Formation with a Lagrangian Neural Cellular Automaton

Cooper Jacobus[1,2] **Beatriz Tucci**[2,3] **Oliver Philcox**[1,2,3]

[1]Department of Physics, Stanford University
[2]Kavli Institute for Particle Astrophysics and Cosmology, Stanford
[3]Leinweber Institute for Theoretical Physics, Stanford

Abstract

To infer the initial conditions of the Universe, we require a model that can accurately describe the non-linear physics of galaxy formation, whilst remaining computationally efficient and fully differentiable. We introduce the *Lagrangian Neural Cellular Automaton* (LNCA), a hybrid deep-learning framework that emulates formation of cosmological structure as a local, iterative dynamical process on a fixed lattice. The LNCA operates in the Lagrangian frame, holding neighborhood connectivity fixed while allowing particle positions to advect freely, replacing expensive scatter-gather graph operations with efficient 3D convolutions. By learning only the *residual* displacement corrections to analytic linear theory (the Zeldovich Approximation), the model dedicates its full expressive capacity to non-linear small-scale clustering. Rotational equivariance is enforced exactly via an integrity-basis construction, and the update rule is parameterized by a compact Kolmogorov–Arnold Network (KAN) for interpretable, sparse dynamics. Trained on the Quijote Latin-hypercube of realistic Cosmological Simulations, the LNCA accurately recovers non-linear displacement fields across a range of times ($z \in \{3, 2, 1, 0.5, 0\}$) and cosmological expansion histories, offering a fast, differentiable forward model suitable for Bayesian field-level reconstruction.

1 Introduction

Next-generation galaxy surveys (DESI [Collaboration et al., 2016], Euclid [Laureijs et al., 2011], LSST [Collaboration et al., 2009]) motivate a transition from summary-statistic compression toward *field-level inference*. In this new paradigm, we aim to jointly reconstruct the initial conditions and the non-linear evolved field, along with the global cosmological parameters, to match observed galaxy data [Jasche and Wandelt, 2013, Wang et al., 2014, Seljak et al., 2017]. The desired forward model should simultaneously satisfy three competing demands: physical accuracy in the non-linear regime ($k \gtrsim 0.1\, h\,\mathrm{Mpc}^{-1}$), computational tractability for millions of sampling steps, and full differentiability for gradient-based optimization [Modi et al., 2021].

Standard N-body simulations satisfy the first criterion but not the latter two. Lagrangian Perturbation Theory (LPT) [Buchert, 1992, Bouchet et al., 1995] is fast and differentiable but breaks down wherever particle trajectories cross. The COLA method [Tassev et al., 2013] provides an effective middle ground by solving large-scale displacements analytically and integrating only the non-linear residual, but relies on particle-mesh force calculations that are costly to differentiate. Neural emulators based on 3D U-Nets [Jamieson et al., 2023, He et al., 2019, Ramanah et al., 2019] achieve high accuracy but operate via single feed-forward passes, lacking temporal continuity. Graph Neural Networks [Sanchez-Gonzalez et al., 2020] respect Lagrangian physics but are computationally prohibitive at cosmological volumes.

We propose the LNCA to bridge these gaps: it inherits the parallel efficiency of Eulerian grid-based computation while retaining the dynamic expressibility of particle-tracking like N-body or graph-based methods, and it integrates continuously in time to produce lightcone-compatible trajectories.

2 Model Design

2.1 Lagrangian Cellular Automaton Structure

The LNCA represents the dark matter fluid as a regular Cartesian lattice of Lagrangian coordinates $\mathbf{q}$. Each node (particle) i is permanently connected to its 26 Moore neighbors in $\mathbf{q}$-space, *regardless* of where it moves in Eulerian space $\mathbf{x}(t)$. Fixed topology eliminates dynamic graph reconstruction and allows all local interactions to be expressed as 3D convolutions in $\mathbf{q}$-space, restoring the speed of grid-based NCAs while preserving continuous particle tracking.

The state of node i at time t is:

$$\bar{S}_i(t) = \left\{ \Psi_i(t),\ \dot{\Psi}_i(t),\ \mathcal{S}_i(t),\ \mathcal{V}_i(t),\ \mathcal{T}_i(t) \right\}, \tag{1}$$

where $\Psi_i = x_i(t) - q_i$ is the displacement vector of particle i, its derivative $\dot{\Psi}_i$ the comoving velocity, and $\mathcal{S}_i \in \mathbb{R}^{L_s}$, $\mathcal{V}_i \in \mathbb{R}^{L_v \times 3}$, $\mathcal{T}_i \in \mathbb{R}^{L_t \times 3 \times 3}$ are latent scalar, vector, and symmetric rank-2 tensor channels that carry memory of the local deformation history.

2.2 Equivariant Update Rule

To guarantee $E(3)$ equivariance, the update network does not operate on raw state vectors. Instead, fixed finite-difference operators compute local gradients, Laplacians, and Hessians over the $\mathbf{q}$-lattice, from which a large family of equivariant terms are extracted. From these fields and the latent states, three feature pools are assembled at each node:

(i) a scalar pool $\mathcal{I}_i$ of $E(3)$-invariant contractions (traces, dot products, tensor–vector products) augmented by the global cosmological parameter vector $\boldsymbol{\theta}_c = \{\Omega_m, \Omega_b, h, n_s, \sigma_8\}$, which describes the gravitational evolution.

(ii) a vector menu $\mathbf{M}_i^{(1)}$ of equivariant vectors (velocity, Laplacians, curls, tensor–vector products).

(iii) a tensor menu $\mathbf{M}_i^{(2)}$ of symmetric rank-2 tensors.

A compact TaylorKAN [Liu et al., 2025a,b] acts point-wise on the invariant scalar pool to produce output weight coefficients $\{\dot{\mathcal{S}}_i, \mathbf{W}_i^{(1)}, \mathbf{W}_i^{(2)}\} = \mathcal{F}_\theta(\mathcal{I}_i)$. Field updates are then synthesized as equivariant linear combinations:

$$\left\{\ddot{\Psi}_i,\ \dot{\mathcal{V}}_i\right\} = \sum_k \mathbf{W}_{i,k}^{(1)} \mathbf{M}_{i,k}^{(1)}, \qquad \dot{\mathcal{T}}_i = \sum_k \mathbf{W}_{i,k}^{(2)} \mathbf{M}_{i,k}^{(2)}. \tag{2}$$

Because the scalar coefficients are invariant and the menus are strictly equivariant, all outputs are mathematically guaranteed to be $E(3)$-equivariant. This choice of architectural restriction is motivated by the success of LPT to capture large scale dynamics with equivariant interactions.

2.3 Integration and Temporal Continuity

Following [Tassev et al., 2013], the total displacement is decomposed as $\Psi(t) = D(t)\,\Psi_{\rm ZA} + \Psi_{\rm res}(t)$, where $D(t)$ is the linear growth factor and $\Psi_{\rm ZA}$ the initial Zeldovich field. The network predicts only the residual acceleration $\ddot{\Psi}_{\rm res}$, hardcoding the large-scale linear dynamics, which are exact, and allowing the network to focus on the non-linear behavior. State variables are advanced via differentiable Euler integration over scale-factor steps Δt:

$$\begin{aligned} \dot{\Psi}_{i,\rm res}(t{+}\Delta t) &= \dot{\Psi}_{i,\rm res}(t) + \ddot{\Psi}_{i,\rm res}(t)\,\Delta t, \\ \Psi_{i,\rm res}(t{+}\Delta t) &= \Psi_{i,\rm res}(t) + \dot{\Psi}_{i,\rm res}(t{+}\Delta t)\,\Delta t. \end{aligned} \tag{3}$$

Latent fields $\mathcal{S}$, $\mathcal{V}$, $\mathcal{T}$ evolve analogously. Because integration is temporally continuous, the LNCA naturally supports the generation of "lightcones" where the simulation's local timeframe is defined relative to the position of an observer. Fig. 1 illustrates the LNCA's dynamic evolution across scale factors.

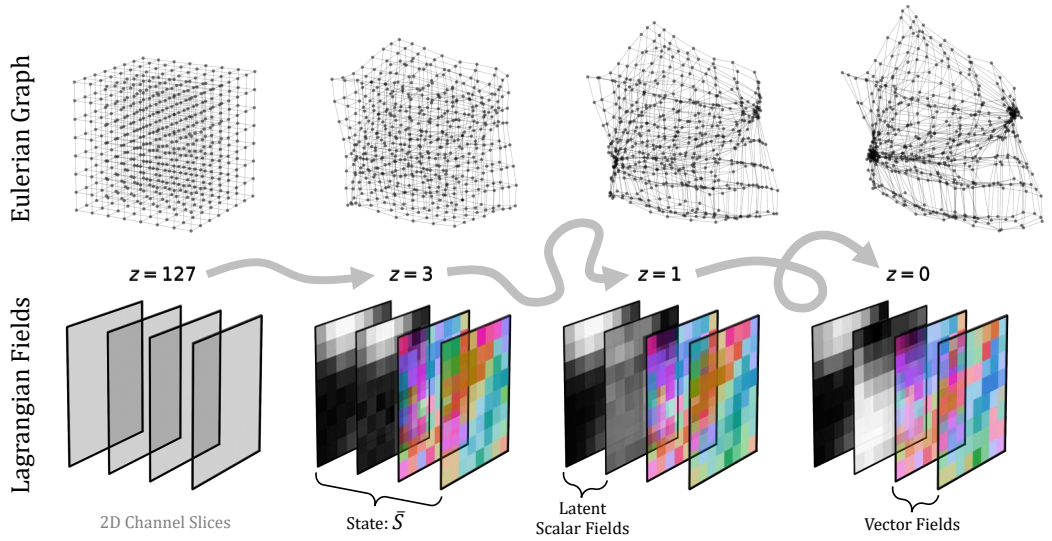

Figure 1: Dynamic evolution of the LNCA forward model. **Top row:** particle positions in comoving Eulerian coordinates at successive scale factors, with edges connecting fixed Lagrangian nearest neighbors. **Bottom row:** Lagrangian latent-field channel maps along a central volume slice at the same times. Non-linear structure emerges from the iterative interaction of these learned fields across many substeps.

3 Training

Data. We use the Quijote Latin-hypercube set [Villaescusa-Navarro et al., 2020], which spans a broad prior over ΛCDM parameters $\{\Omega_m, \Omega_b, h, n_s, \sigma_8\}$. The LNCA is initialized at $z_0 = 127$ with the linear displacement field and integrated to $z = 0$. Loss is evaluated at $z \in \{3, 2, 1, 0.5, 0\}$ via backpropagation through time (BPTT). A spatial padding strategy evaluates dynamics on a 64^3 volume and crops a 16-voxel boundary, restricting gradients to the uncontaminated 32^3 core.

Loss function. The physics loss combines a normalized Lagrangian displacement term, $\mathcal{L}_{\rm disp}$ defined as the normalized $\mathcal{L}_2$ distance between the true and predicted displacement vector fields:

$$\mathcal{L}_{\rm disp} = \frac{\left\langle \|\Psi_{\rm pred} - \Psi_{\rm true}\|^2 \right\rangle}{\left\langle \|\Psi_{\rm true} - \Psi_{\rm ZA}\|^2 \right\rangle + \varepsilon}, \tag{4}$$

with an Eulerian power spectrum residual $\mathcal{L}_{\rm pk}$ defined analogously. Values below unity indicate improvement over linear theory. Preliminary models plateau around $\mathcal{L}_{\rm disp} \simeq \mathcal{L}_{\rm pk} \simeq 0.3$, or an improvement of about 70% in these objectives.

A hierarchical sparsity penalty $\mathcal{L}_{\rm sparse} = \lambda\big[\sum_f \|\mathbf{W}_{1,f}\|_2 + \sum_{p=2}^{3} \frac{p}{2}\|\mathcal{C}^{(p)}\|_F\big]$ applies progressively stronger Occam priors to higher polynomial degrees, encouraging the model to prune any unused or under-contributing pathways as training continues.

Multi-generational evolutionary training. Standard gradient descent tends toward a single locally sparse solution. We embed training within a multi-generational evolutionary outer loop over a population of $N_{\rm pop} = 8$ models across $N_{\rm gen} = 8$ generations. Each generation comprises three phases: (i) *Growth* (moderate sparsity, $\lambda = 0.1$, 1000 steps); (ii) *Competition* (aggressive sparsity $\lambda = 10$, magnitude pruning, 1000 steps); and (iii) *Selection* via NSGA-II non-dominated sorting on the objective vector $(\langle\mathcal{L}_{\rm disp}\rangle, \langle\mathcal{L}_{\rm pk}\rangle, \mathcal{L}_{\rm sparse})$, with species-preserving diversity control. Survivors breed via structure-aware crossover; extinct pathways undergo neurogenesis ($\sigma_{\rm ng} = 10^{-5}$) to regenerate structural diversity. The balance between the Growth and Competition phases corresponds to the competing needs for the models to freely explore parameter space and ruthlessly exploit it; experimentation with this balance is still underway.

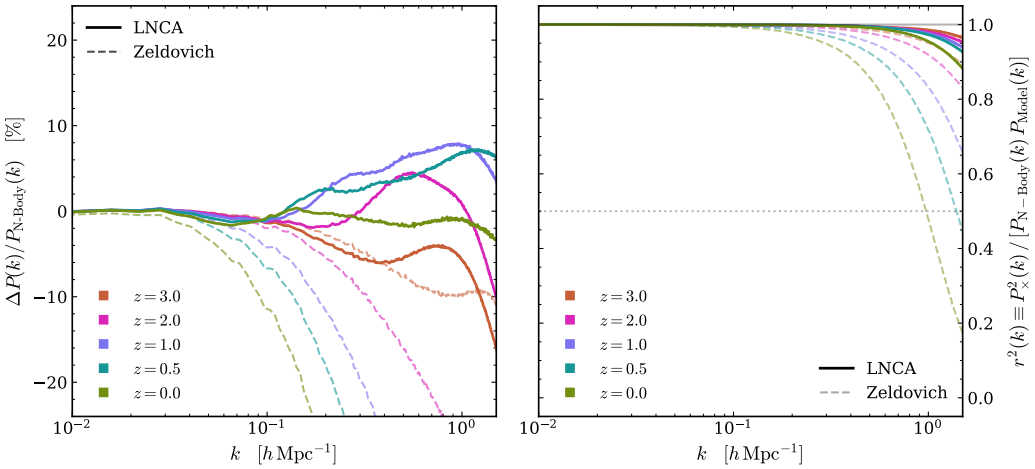

Figure 2: Summary statistic results of our model. **Left:** the spherically-average 1D power spectrum produced by the Zeldovich Approximation and our Model compared to the true N-body simulated power spectrum. **Right:** Cross-spectrum of the ZA and our model with the target field.

4 Validation Results and Discussion

We evaluate our model against the Zeldovich Approximation (ZA), which serves as the physics-informed lower bound. Performance is assessed via the standard matter power spectrum $P(k)$ and the cross-correlation $r(k)$ between predicted and ground-truth N-body Eulerian density fields.

Power Spectrum Accuracy. Our model achieves consistent improvement over the ZA across the mildly non-linear regime. At $k \lesssim 0.3\,h/\mathrm{Mpc}$, both the ZA and our model recover the linear power spectrum faithfully, as expected. The gains become appreciable at $0.3 \lesssim k \lesssim 1.0\,h/\mathrm{Mpc}$, where our model reduces the fractional power spectrum error relative to the N-body ground truth by approximately 35–50% compared to the ZA, recovering non-linear mode coupling that first-order Lagrangian perturbation theory systematically misses.

Cross-Correlation Coefficient. The cross-correlation coefficient $r(k)$ between the predicted and true density fields provides a more stringent test of phase-level fidelity. The ZA maintains $r(k) \gtrsim 0.95$ only for $k \lesssim 0.1\,h/\mathrm{Mpc}$, degrading significantly at smaller scales where non-linear structure formation scrambles the phase information. Our model extends this high-correlation roughly 50% deeper into the non-linear regime.

Comparison to ML Baselines. Relative to comparable ML emulators [Jamieson et al., 2023], our model offers nearly competitive summary-statistic performance while providing substantially greater parameter efficiency (> 100 times fewer terms) and architectural transparency. Feed-forward volumetric architectures achieve marginally lower power spectrum residuals, but do so via opaque global mappings with no correspondence to the underlying gravitational dynamics. By contrast, our model's latent field evolution can be directly inspected and correlated with physical quantities.

Conclusion. We have introduced our model, a cosmological emulator that couples a Lagrangian Neural Cellular Automaton with Kolmogorov-Arnold Network operators to produce differentiable, temporally continuous approximations to N-body displacement fields. The model achieves moderate but consistent improvement over the Zeldovich Approximation in both power spectrum fidelity and phase-level cross-correlation across the mildly non-linear regime, while offering substantial improvement in interpretability and temporal coherence over comparable ML architectures. The decomposition of latent dynamics provides a transparent window into the learned approximation that purely data-driven approaches cannot offer. Future work will refine the model's accuracy and integrate the emulator as a differentiable forward model of the likelihood component within a full field-level inference pipeline.

References

DESI Collaboration et al. The desi experiment part i: Science,targeting, and survey design. *arXiv preprint arXiv:1611.00036*, 2016.

R Laureijs et al. Euclid definition study report. *arXiv preprint arXiv:1110.3193*, 2011.

LSST Science Collaboration et al. Lsst science book, version 2.0. *arXiv preprint arXiv:0912.0201*, 2009.

Jens Jasche and Benjamin D Wandelt. Bayesian physical reconstruction of initial conditions from large scale structure surveys. *Monthly Notices of the Royal Astronomical Society*, 432(2):894–913, 2013.

Huiyuan Wang, HJ Mo, Xiaohu Yang, YP Jing, and WP Lin. Elucid-exploring the local universe with the reconstructed initial density field. i. hamiltonian markov chain monte carlo method with particle mesh dynamics. *The Astrophysical Journal*, 794(1):94, 2014.

Uroš Seljak, Grigor Aslanyan, Yu Feng, and Chirag Modi. Generative model for cosmological large-scale structure. *Journal of Cosmology and Astroparticle Physics*, 2017(12):009, 2017.

Chirag Modi, Yu Feng, and Uroš Seljak. Cosmological inference with differentiable forward modeling. *Journal of Cosmology and Astroparticle Physics*, 2021(10):035, 2021.

Thomas Buchert. Lagrangian theory of gravitational instability of friedman-lemaitre cosmologies and the'zel'dovich approximation'. *Monthly Notices of the Royal Astronomical Society*, 254(4): 729–737, 1992.

F Bouchet, S Colombi, E Hivon, and R Juszkiewicz. Perturbative lagrangian approach to gravitational instability. *Astronomy and Astrophysics*, 296:575, 1995.

Svetlin Tassev, Matias Zaldarriaga, and Daniel Eisenstein. Cola: fast cosmological simulations using the lagrangian perturbation theory with a particle-mesh algorithm. *Journal of Cosmology and Astroparticle Physics*, 2013(06):036, 2013.

Drew Jamieson, Chuangyin Yin, Yu Feng, and Uroš Seljak. Field-level neural network emulator for cosmological n-body simulations. *Physical Review D*, 108(12):123514, 2023.

Siyu He, Yin Li, Yu Feng, Shirley Ho, Siamak Ravanbakhsh, Wei Chen, and Barnabás Póczos. Learning to predict the cosmological structure formation. *Proceedings of the National Academy of Sciences*, 116(28):13825–13832, 2019.

Doogesh Kodi Ramanah, Tom Charnock, and Guilhem Lavaux. Painting halos from 3d dark matter fields using generative adversarial networks. *Astronomy & Astrophysics*, 621:A69, 2019.

Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter Battaglia. Learning to simulate complex physics with graph networks. *International conference on machine learning*, pages 8459–8468, 2020.

Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić, Thomas Y. Hou, and Max Tegmark. Kan: Kolmogorov-arnold networks, 2025a. URL `https://arxiv.org/abs/2404.19756`.

Ziming Liu, Max Tegmark, Pingchuan Ma, Wojciech Matusik, and Yixuan Wang. Kolmogorov-arnold networks meet science. *Phys. Rev. X*, 15:041051, Dec 2025b. doi: 10.1103/4t7t-v19l. URL `https://link.aps.org/doi/10.1103/4t7t-v19l`.

Francisco Villaescusa-Navarro, ChangHoon Hahn, Elena Massara, Arka Banerjee, Ana Maria Delgado, Doogesh Kodi Ramanah, Tom Charnock, Elena Giusarma, Yin Li, Erwan Allys, Antoine Brochard, Cora Uhlemann, Chi-Ting Chiang, Siyu He, Alice Pisani, Andrej Obuljen, Yu Feng, Emanuele Castorina, Gabriella Contardo, Christina D. Kreisch, Andrina Nicola, Justin Alsing, Roman Scoccimarro, Licia Verde, Matteo Viel, Shirley Ho, Stephane Mallat, Benjamin Wandelt, and David N. Spergel. The quijote simulations. *The Astrophysical Journal Supplement Series*, 250(1):2, aug 2020. doi: 10.3847/1538-4365/ab9d82. URL `https://doi.org/10.3847/1538-4365/ab9d82`.