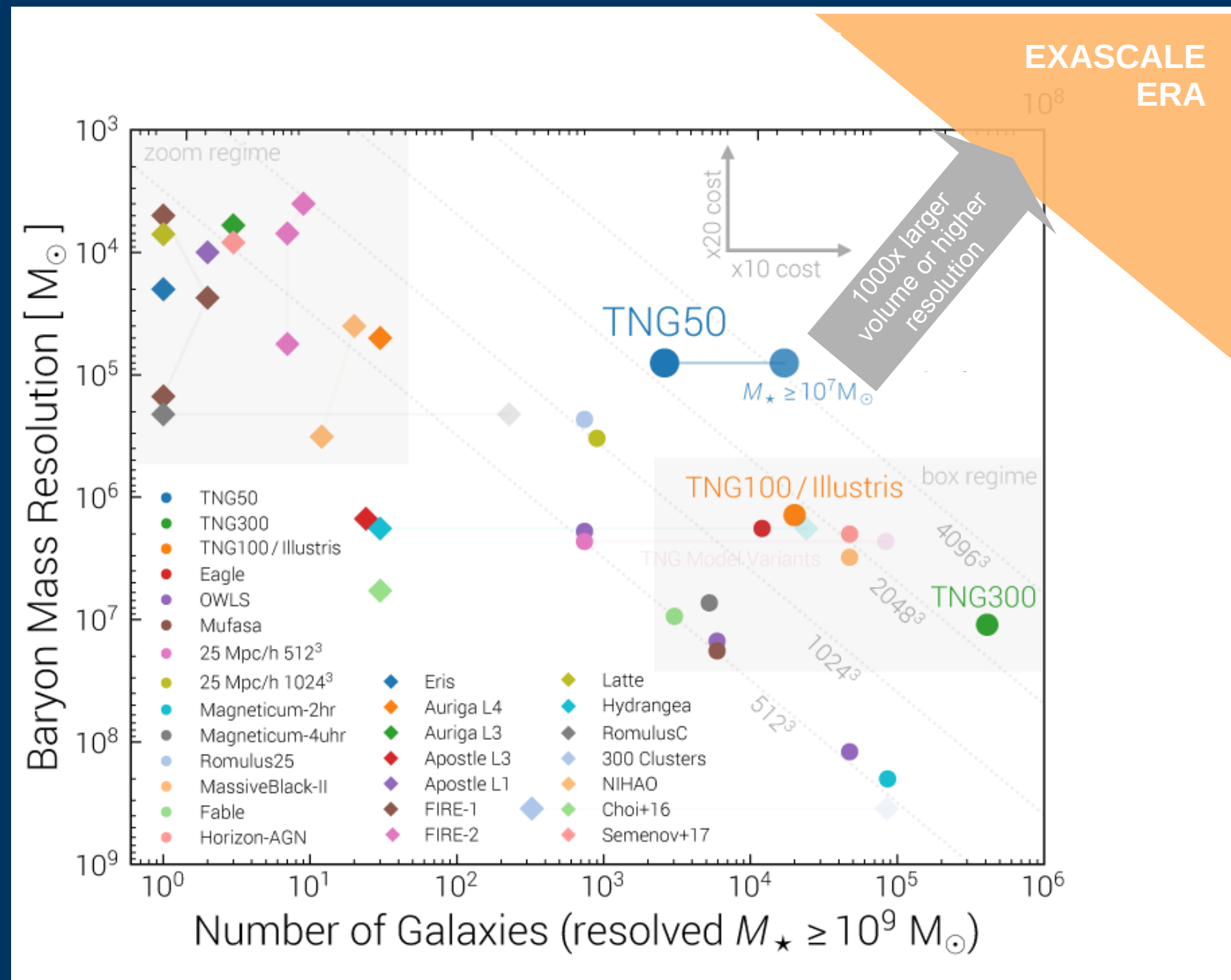


Knowledge discovery in simulations using Representation Learning

- Exploration and analysis of large, high-dimensional datasets from *any cosmological simulation* in preparation for Exascale era
- Similarity space based on the intrinsic distribution of the data
- Browser demo available at space.h-its.org



Sebastian T. Gomez (HITS, Heidelberg)

Building Galaxies from Scratch / 22.2.2024 / Vienna

Representation Learning for Knowledge Discovery in Cosmological Simulations

Sebastian Trujillo-Gomez, Kai Polsterer, Bernd Doser, & Andreas Fehner

1. The Challenge

- Simulations are the best approximation to experimental laboratories in Cosmology. However, the size and complexity of their outputs makes them difficult to explore, visualize, and analyze, and limits the interpretability of their predictions
- Simulations present very different data challenges compared to observations. They now follow the evolution of a universe in a box including $\sim 10^9$ particles and $\sim 10^4$ galaxies in a 2D-dimensional feature space ~ 1 PB of data
- Today's largest projects (e.g. EAGLE, TNG, Magneticum, Millennium, FLAMINGO) are already difficult to explore and analyze, and this will become an even greater challenge for Exascale simulations
- We need to leverage unsupervised Deep Learning to efficiently extract information and generate knowledge from these huge datasets

2. Our Solution

- We are developing a software tool that
 - agnostically, automatically, and efficiently learns a model for the underlying structures in the simulation using arbitrary fields, and then reduces the dimensionality of the data by projecting it onto a simple 2D latent space
 - provides an intuitive interpretation of the low-dimensional space by projecting both the dataset and the representation onto a hierarchical sphere for easy interactive inspection, sample selection, and downstream quantitative analysis

The diagram illustrates the pipeline: 'Automatically process outputs to select object type (e.g. galaxies) and fields of interest (e.g. 2D or 3D maps)' leads to a 'Generative Deep Learning Model (e.g. Convolutional Autoencoder)'. This model performs 'compact representation' of the data. The resulting representation is used for 'projection of data onto sphere', creating a 'hierarchical model' (a sphere of points). This sphere is used for 'catalog creation' and 'data selection and retrieval based on structural similarity'. The sphere also enables 'inspection of detailed properties of individual objects' and 'data projection' for 'local data analysis' using tools like TOPCAT and ALADIN.

3. Implementation

- We use a Generative Deep Learning model (a Hyperspherical Variational Convolutional Autoencoder; Davidson et al. 2018) to learn a compact 2D representation of the simulated structures
- The data and the model latent space are mapped onto a hierarchical spherical HPS³-tiling for visualization
- The projection can be explored interactively to visually examine the structures at each level of the hierarchy, and to select samples of objects that are output directly to 'Topcat' tables for local analysis

4. What can I use it for?

Some examples:

- Learn and visualize a structural similarity space of simulated structures in an arbitrary number of features for any matter component (e.g. stars, gas and dark matter)
- Easily explore relationships between different properties of the structures (e.g. galaxies, large-scale environment) by projecting additional features onto the representation space
- Compare directly the distributions of simulations and next-generation survey data in the same representation space

Do you have simulations with large complex outputs? Contact us at sebastian.trujillo@h-its.org

References: Davidson T. R., Fehner A., De Cao N., Klot T., Tomczak J. M., 2018, doi:10.48550/arXiv.1804.00991; Nelson D., Springel V., Pillepich A., Rodriguez-Gomez V., Tacchi F., Genel S., Vogelsberger M., et al., 2019, *Comput. & Appl. Math.*, 6, 2, doi:10.1007/s40348-019-0026-4; Rodriguez-Gomez V., Snyder G. F., Lotz J. M., Nelson D., Pillepich A., Springel V., Genel S., et al., 2018, *MNRAS*, 481, 4340, doi:10.1093/mnras/sty3345; Taylor M. B., 2005, *ASPC*, 342, 29 (Topcat)

Acknowledgements: Funded by the European Union. This work has received funding from the European High-Performance Computing Joint Undertaking (JU) and Belgium, Czech Republic, France, Germany, Greece, Italy, Norway, and Spain under grant agreement No101019718. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European High-Performance Computing Joint Undertaking (JU) and Belgium, Czech Republic, France, Germany, Greece, Italy, Norway, and Spain. Neither the European Union nor the granting authority can be held responsible for them. We gratefully acknowledge the generous and invaluable support of the Klaus Tschira Foundation. This research has made use of 'ALADIN sky atlas' developed at CDS, Strasbourg Observatory, France. We are grateful to Thomas Behr for help with ALADIN, and to Dylan Nelson and the 'TNG Collaboration' for providing access to their data and visualizations.

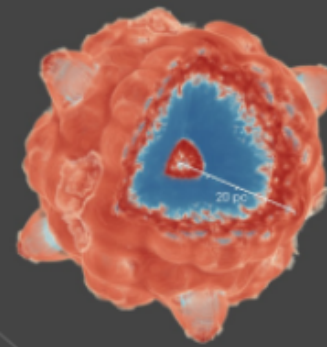
Cloud Formation by Supernova Implosion

Leonard Romano, Manuel Behrendt, Andreas Burkert

New characterization of

radiative stage &

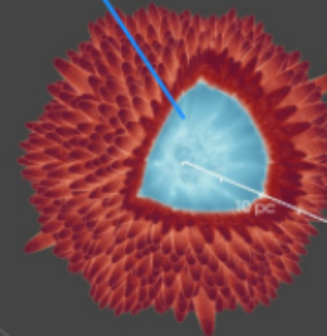
merging with ISM



d) Cloud formation
($t \gtrsim \text{few Myr}$)

$$M_{\text{cloud}} \sim 10^3 - 10^4 M_{\odot}$$

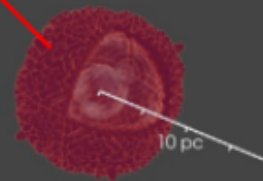
Inflowing Gas



c) Implosion
($t \lesssim \text{few Myr}$)

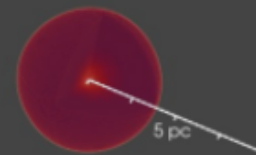
$$P_{\text{shell}} \sim P_{\text{ISM}}$$
$$\dot{M}_{\text{in}} > 0$$

Outflowing Gas



b) Momentum Conserving Snowplow
($t \lesssim 20 - 100 t_{\text{sf}}$)

$$P_{\text{shell}} \gg P_{\text{ISM}}, P_{\text{Bubble}}$$
$$M_{\text{Bubble}} \sim 0$$



a) Pressure Driven Snowplow
($t < 3 - 5 t_{\text{sf}}$)

$$P_{\text{Bubble}} \gg P_{\text{shell}} \gg P_{\text{ISM}}$$
$$\dot{M}_{\text{Bubble}} < 0$$



Scan QR-code to view poster now!