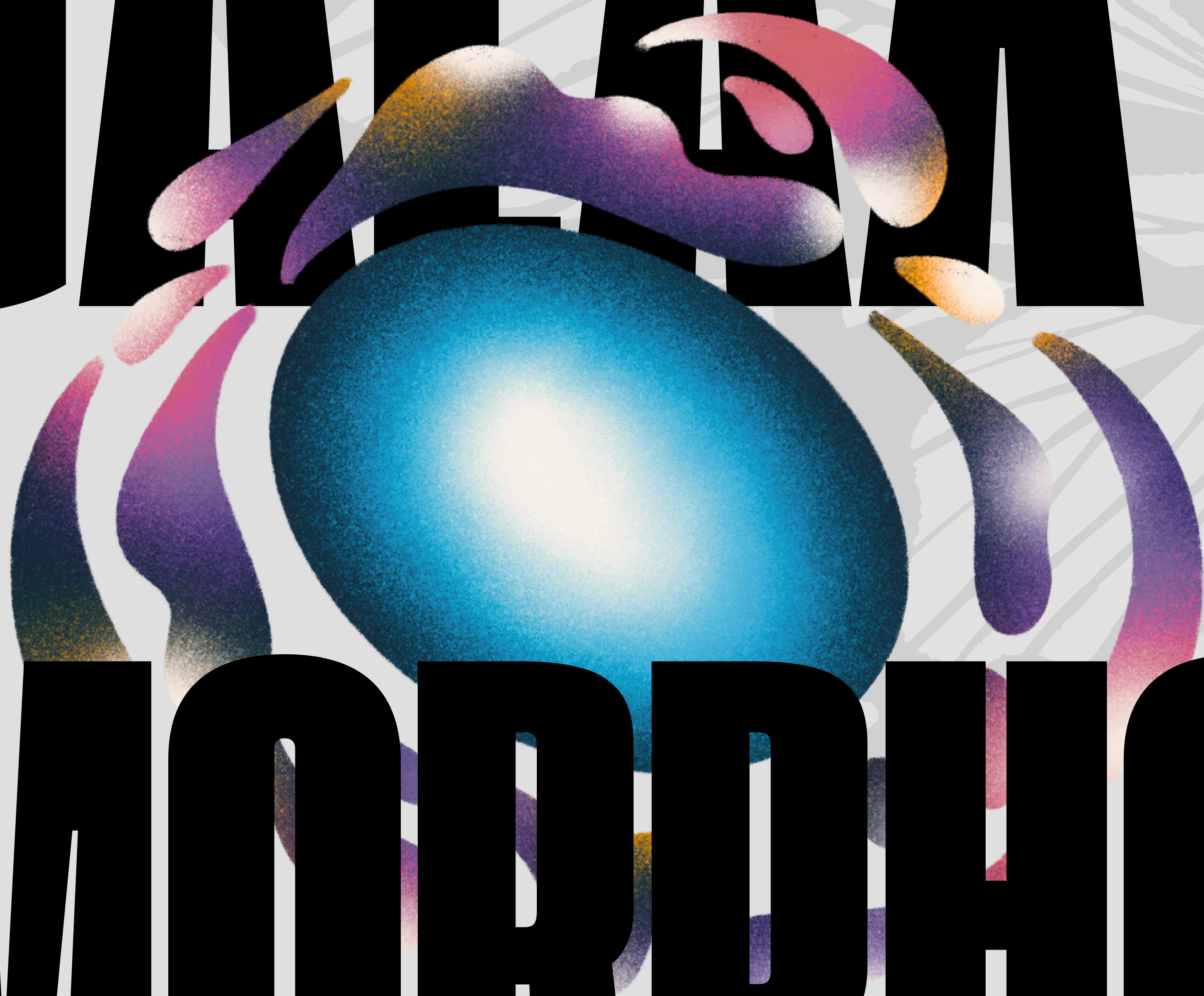
ESTD 13 BILLION YEARS AGO

GALAXY MORPHO



EQUINOX

SUMMER RESEARCH MODULE

SPIRAL **ELLIPTICAL**

Explore the physics behind galaxy formation and apply machine learning to classify real galactic structures. A fusion of cosmological theory and data-driven modeling to decode the universe.

Introduction

Galaxies are the luminous building blocks of the cosmos — vast systems of stars, gas, dust, and dark matter that stretch across millions of light-years. **Their origins trace back to the early universe**, when tiny quantum fluctuations evolved into the large-scale structure we observe today. Over billions of years, processes like gravitational collapse, angular momentum conservation, star formation, supernova feedback, and mergers have shaped galaxies into the diverse forms we see — **from elegant spirals to massive ellipticals and chaotic irregulars.**

Understanding how galaxies form, evolve, and acquire their morphology is central to modern astrophysics. Equally important is the ability to analyze and interpret the immense datasets produced by contemporary sky surveys. This project combines both: **a study of the physical principles that govern galactic structure and dynamics, alongside computational tools that leverage real data to classify galaxies based on observable features.**

By engaging with both the theoretical foundations and data-driven methodologies, this project offers a holistic exploration of the cosmos — one that connects the forces that shape galaxies with the algorithms we use to study them. **It invites you to think like a physicist, work like a data scientist, and discover like an astronomer.**

Whether you're driven by curiosity about the cosmos or inspired by the power of data-driven discovery, **it invites you to dive deep into the interplay between physics and computation**. Explore boldly, think critically, and **enjoy the process of unraveling the story of galaxies.**

Module One — 40 points

Task 1: Comprehensively describe the formation of galaxies from the early universe to present day.

Your response should address the following key aspects:

- **Dark Matter Framework:** Explain the role of dark matter halos and how primordial density fluctuations evolved into the cosmic web structure we observe today
- **Baryonic Physics:** Describe gas cooling processes in brief, condensation within dark matter halos, and the mechanisms that trigger star formation in early galaxies
- **Stellar Evolution:** Detail the transition from the first Population III stars to Population II stars and their impact on galaxy chemical evolution
- **Feedback Processes:** Explain how stellar winds, supernovae explosions, and active galactic nuclei (AGN) regulate star formation and shape galaxy evolution
- **Hierarchical Assembly:** Describe the bottom-up formation process through major and minor galaxy mergers and how this affects final galaxy morphology
- **Environmental Influences:** Discuss how different environments (galaxy clusters vs. field galaxies) affect galaxy formation and morphological development

Include appropriate diagrams or sketches to illustrate key concepts.

Task 2: Using Newtonian physics, mathematically demonstrate why galaxies tend to flatten into disk structures.

Provide a complete mathematical analysis that caters to the following points of discussion:

Module One

Part A - Angular Momentum Conservation

- Derive the relationship between angular momentum and disk formation.
- Show mathematically why a spherical gas cloud with net rotation will collapse preferentially along the rotation axis.
- Use the specific angular momentum expression: $\boldsymbol{j = r \times v}$

Part B - Energy Considerations

- Apply the **virial theorem** to demonstrate energy redistribution during gravitational collapse.
- Show how gravitational potential energy converts to kinetic energy differently in radial vs. tangential directions
- Derive the condition for rotational support against gravitational collapse.

Part C - Force Balance Analysis

- Demonstrate the balance between gravitational and centrifugal forces.
- Show that: $\boldsymbol{GMm/r^2 = mv^2/r}$ for circular orbits.
- Calculate the critical rotation rate required for disk formation.

Part D - Numerical Application

- Use typical galaxy parameters: **Mass (M) = $10^{11}\ M_\odot$, Radius (R) = 10 kpc**
- Provide step-by-step calculations with these values.

Include at least one force-balance diagram for visual depiction.

Connecting the dots

In the previous module, we studied the **physical principles that govern the formation and morphological evolution of galaxies**—ranging from hierarchical mergers and angular momentum conservation to the influence of feedback and environment. In this module, you will apply these theoretical insights to a modern data-driven challenge: the **automated classification of galaxy morphology** using real observational data and supervised machine learning algorithms.

Morphological classification of galaxies (e.g., elliptical, spiral) is essential for understanding galaxy evolution, stellar populations, and the large-scale structure of the universe. Traditionally, morphology has been determined visually by astronomers or through citizen science efforts (e.g., Galaxy Zoo). However, with the increasing volume of imaging data from modern surveys such as the Hyper Suprime-Cam (HSC) and upcoming missions like LSST and Euclid, manual classification is no longer feasible. This motivates the need for **robust, scalable machine learning techniques that can classify galaxies based on their photometric and structural features.**

In this module, you will develop a **supervised learning pipeline** to classify galaxies into two primary morphological categories—**spiral and elliptical**—using features derived from the **GalaxiesML dataset**. This dataset contains over 280,000 galaxies with five-band photometry (g, r, i, z, y), redshifts, and morphological parameters (e.g., Sérsic index, half-light radius, ellipticity) extracted from **real imaging data**. You will define **morphology classes based on physical indicators** such as the Sérsic index and train models to learn this classification from the remaining tabular features.

The goal is not only to achieve **high classification accuracy but also to interpret the physical meaning of the learned features** and compare your machine-derived classifications to known theoretical expectations from Module 1.

Module Two — 40 points

Task 1: Explore and define a method for labeling galaxy morphologies.

- Examine the dataset to identify structural parameters—such as the Sérsic index, ellipticity, or size indicators—that could serve as proxies for galaxy morphology. Use your understanding of galactic structure to formulate a binary classification scheme (e.g., elliptical vs. non-elliptical). Justify your chosen approach using trends in the data or physical reasoning.

Task 2: Investigate and preprocess the dataset for machine learning.

- Delve into the dataset and analyze the nature of its features. Consider how to handle missing or uncertain values, normalize numerical parameters, and possibly engineer new features that capture astrophysical insights. Reflect on which features may be most informative for distinguishing galaxy types.

Task 3: Design and implement machine learning models for classification.

- Select at least two supervised learning algorithms—such as but not limited to k-Nearest Neighbors, Support Vector Machines, or Random Forests—and train them using your labeled data. You are encouraged to explore different models, experiment with their parameters, and observe how they perform under various settings.

Task 4: Evaluate and compare your models.

- Use suitable performance metrics — like F1-score, confusion matrices and others — to assess the effectiveness of each model. Interpret the results not only in terms of performance but also in relation to the nature of the galaxy data. What strengths or limitations do you observe in each approach?

Module Two

Task 5: Analyze the most influential features in your models.

- Explore which features contributed most to the classification in each model. What do these features tell you about the physical differences between galaxy types? Use your findings to draw connections between data-driven results and underlying astrophysical principles.

The **dataset and metadata** along with their descriptions, for the above discussed analysis, can be found at the given link: **GalaxiesML**

Having explored the physical mechanisms behind galaxy formation and applied supervised learning techniques to classify real galactic data, you now possess a multifaceted understanding of how galaxies evolve and how we study them.

As an **optional extension**, you are encouraged to build a **simple simulation** in the next module that models **galaxy formation** from first principles. This **bonus module** offers a dynamic and visual perspective into how gravity, angular momentum, and energy conservation shape galactic structure — reinforcing the concepts covered so far while introducing computational physics techniques.

This exploratory task is designed for those eager to dive deeper into the physics of structure formation and test their coding and modeling skills. While a bit challenging, it can be immensely rewarding and offers a glimpse into how theoretical astrophysics meets numerical experimentation.

Module Three — 20 points

Task 1: Construct a 2D simulation framework for galaxy formation

- Build a particle system that includes both dark matter and gas particles, each characterized by attributes like position, mass, and velocity. Implement gravitational interactions using Newton's law of universal gravitation, and apply softening techniques to avoid numerical divergences in close particle encounters. Choose a suitable numerical integration method (e.g., Euler or Runge-Kutta) to evolve the system over time.

Task 2: Implement physical processes within the simulation

- Incorporate the key physical mechanisms behind galaxy formation into your model. Simulate gravitational collapse, enforce angular momentum conservation, and track changes in energy. Ensure your simulation can monitor total energy, kinetic energy, and potential energy over time to validate physical consistency.

Task 3: Visualize and analyze system evolution

- Create dynamic visualizations of the evolving galaxy. Generate an animated plot of particle positions over time to visualize structure formation. Additionally, plot time series of total energy and angular momentum to verify conservation and observe emergent behavior.

Task 4: Experiment with different initial conditions

Run simulations under three different initial angular momentum configurations:

- Case A: High angular momentum → spiral galaxy structure
- Case B: Low angular momentum → elliptical-like collapse
- Case C: Intermediate angular momentum with asymmetry → irregular formation

Compare the outcomes across these scenarios and relate them to theoretical expectations about galaxy morphology and dynamics.

Guidelines

Report Submission

- Format: Results from all modules (Module 1, Module 2, and optional Module 3) should be compiled into a well-organized report.
- Acceptable formats: PDF generated via Word, LaTeX or other editors.
- Content:
  - All theoretical discussions, derivations, and explanations.
  - Visuals and figures (e.g., diagrams, plots etc.).
  - Summary of model performance (for classification module).
  - Brief conclusion reflecting key learnings.
  - If using external datasets or tools, clearly cite the source.

Code Submission

- Format: All code must be submitted in Jupyter Notebook (.ipynb) format.
- Structure:
  - Separate notebooks for each module (Module 2 and optional Module 3).
  - For Module 3 (if attempted), ensure all animations or outputs are viewable and reproducible.
- Readability:
  - Code should be clean, modular, and well-commented.
  - Include a brief README for running the codes

Submission Format

- Final submission folder should include:
  - Report PDF
  - One or more .ipynb notebooks
  - Any auxiliary files (e.g., images, data extracts, etc.)
- Compress the final folder to be uploaded as a ZIP file named: GalaxyMorpho_<YourTeamName>.zip