Proposal for Project 03

Daniel Villarruel-Yánez & Franklin Limaico

22nd October 2025

1 Problem

Simulations of active matter, while powerful, are computationally intensive, limiting large-scale parameter exploration and real-time analysis. Standard data-driven surrogate models can accelerate these simulations but often fail to respect fundamental physical laws, leading to non-physical artifacts, instability, and poor generalization to new parameter regimes. This project aims to address this gap by developing a hybrid modelling framework that integrates known physical constraints directly into the machine learning architecture. Our goal is to build interpretable, robust, and generalizable surrogate models of active matter dynamics.

2 Approach

We will be using **The Well Dataset** (*active-matter*) for 24 repetitions of 81-frame simulations of active rod-like particles in a 256×256 grid. We will follow this general scheme:

Phase 1. Characterisation of the physics of the system and formulation of constraints.

- Analyse the active matter systems to identify physical constraints and symmetries.
- Each identified constraint/symmetry will be translated into a mathematical expression.
- Map the mathematical expressions to loss terms and architectural constraints.

Phase 2. Development of the baseline model.

- Model architecture design based on pattern recognition from previous frames.
- Evaluation of the model and analysis of failures.

Phase 3. Design of the hybrid, physics-informed model.

- Review existing literature on Physics-Informed Neural Networks.
- Set Physics-based loss.
- Set constraint layers.
- Evaluate advanced metrics, including norm of physical residuals, the divergence of vector fields.

3 Hypotheses

A. The pattern-driven baseline model will exhibit significant drift and generate non-physical artifacts in long-term rollouts, failing to generalise to parameter regimes outside its immediate training distribution.

B. A hybrid model incorporating physical constraints as loss or architecture based constraints will produce more stable, physically plausible predictions and demonstrate superior generalization capabilities.

C. By embedding physical knowledge, the hybrid model will be more data-efficient and robust to inputs compared to the baseline emulator.

4 Integration into the Join Manuscript

This project forms the technical contribution to the joint paper. The baseline model's successes and its possible failures will motivate the need for physics-informed approaches. The design and evaluation of the final hybrid model will serve as the primary methodology and provide the key results, demonstrating how integrating physical principles with machine learning provides powerful and reliable surrogate models for complex active matter systems.