

# 1 Introduction

This honor thesis is motivated by my previous research on building a computer system for simulating and analyzing a regulatory biological network called *zebrafish segmentation clock network*. In particular, the computer system builds on top of a list of species (such as mRNAs and proteins of *her1*, *her7*, *her13* genes) and a set of differential equations describing reactions (such as protein synthesis, protein degradation and mRNA synthesis) in this regulatory network. The system workflow consists of several phases: parameter estimation, simulation, feature extraction, and testing. Parameter estimation is used to generate random parameter values within a certain range through a genetic algorithm in hopes of finding a “desirable” parameter set whose simulation result “fits” the biological model. Simulation, based on a particular parameter set, happens at tissue level (multiple cells) and is responsible for replicating the entire biological process for a certain period of time. This simulation is also tasked with recording concentration levels of each species during this process. Feature extraction retrieves concentration levels recorded in the previous step and turns those values into properties such as period, amplitude, and synchronization score. At last, during testing, the properties produced in the previous step are compared to standards created based on experimental data; this comparison determines if the results are “desirable” and “fit” the model.

## 1.1 Motivation

Various systems and packages have been created to meet the need of research in regulatory biological networks. Some of the created packages provide a user-friendly interface, yet they either conduct simulations on a small-scale basis, or they focus solely on ordinary differential equations, and thus ignoring delay differential equations. Both of the practices effectively prevent researchers from building accurate and comprehensive models. Multiple cell simulation is important in analyzing the effects of intercellular communication on various species in the system, and due to the time delay in certain biological processes, delay differential equations are sometimes crucial in analyzing such network structures. For example, gene expression in zebrafish segmentation clock system consists of two main segments, gene transcription and mRNA translation. Because of the biologically complex nature of gene transcription and mRNA translation, such reactions typically take five to ten minutes to finish in the cell. Thus, the corresponding computer simulation relies on DDEs to mimic the time delay introduced by such processes in a biological system. Other

packages, though never previously applied in this particular field, can offer potentially desirable runtime for large-scale simulations through GPU acceleration. However, since those packages are often presented as libraries, they usually require a decent understand of GPU programming for researchers to enjoy such benefits. Under such circumstances, many researchers are forced to spend prolonged periods of time creating systems to simulate specific models. Such systems generally lack flexibility or configurability, and without GPU acceleration, tissue level simulations running on CPU are far less efficient and extremely time-intensive, even on computer cluster systems. Here, I built a system that is compatible with DDEs, flexible, easily configurable, and accessible to researchers without substantial programming skills.

### 1.2 Benefits

This new system will provide researchers with considerable advantages in three key areas. First, the cost of creating a running system for a comprehensive large-scale model will be significantly lower, and when researchers are able to configure and simulate their own model, more time can now be devoted to alternative research tasks. Second, runtime of the new system shows significant improvement; it is five to ten times faster than CPU only systems. This is important to researchers since parameter estimation usually takes days to complete on computer clusters, and much longer on usual workstations. Third, requirements on hardware for efficient usage of the systems will be greatly reduced. While previous systems usually require clusters for parameter estimations, the new developed system will be compatible with most workstations that contain a powerful GPU card. Compared to a cluster of powerful cores that typically cost 100,000+ USD, a GPU card is economically efficient (typically < 5,000 USD), requires less maintenance, and is widely accessible to researchers.