Project Vision Document

Project Title: GPU-Accelerated Molecular Dynamics for Neurodegenerative Drug Discovery

Team Name: Team Acelot

Team Members:

- Vincent Zhu: vincentzhu@ucsb.edu
- Bryan Huang: b_huang@ucsb.edu
- Kevin Deng: kevindeng@ucsb.edu
- James Fu: fu77@ucsb.edu
- George Nebieridze: george709@ucsb.edu

Team Lead: George Nebieridze

Company Overview:

Acelot is a drug discovery startup focusing on neurodegenerative diseases. Its mission is to develop innovative therapies targeting these diseases, which often pose significant challenges for traditional drug discovery approaches. The company is currently testing a new drug design aimed at addressing key issues with TDP-43, a protein associated with several neurodegenerative conditions.

Problem Statement:

Drug discovery is an expensive, lengthy, and resource-intensive process. For Acelot, each compound synthesis costs approximately $5,000 and takes about two weeks. Due to rigorous synthesis, screening, clinical trials, and regulatory approval processes, bringing a drug to market can require $0.5 to $1 billion and take several years.

One of the significant challenges facing Acelot's drug discovery efforts is the behavior of **TDP-43**, a critical protein involved in neurodegenerative diseases. The accumulation of TDP-43 aggregates in the central nervous system is a common feature of many neurodegenerative diseases, such as amyotrophic lateral sclerosis (ALS), frontotemporal dementia (FTD), Alzheimer's disease (AD), and limbic predominant age-related TDP-43 encephalopathy (LATE). TDP-43 is difficult to crystallize, which complicates its study in drug design. To overcome this, Acelot relies heavily on **molecular dynamics (MD)** simulations to understand the protein's behavior and interactions with potential compounds.

However, traditional molecular dynamics simulations face two key bottlenecks:

1. **Computational Expense**: Full replica molecular dynamics simulations are computationally expensive.
2. **Performance Delays**: Event-based molecular dynamics simulations often take too long to deliver actionable results, slowing down the drug development cycle.

Project Overview:

The goal of this project is to **optimize Acelot's event-based molecular dynamics (MD) code to leverage the power of GPUs** for improved performance and scalability. By optimizing the MD code using CUDA (Compute Unified Device Architecture), the project aims to significantly reduce the time and cost associated with molecular dynamics simulations. Right now, the code uses a single core on a CPU to run the simulation, and we are planning to optimize the code by utilizing 9,000 GPU cores in order to speed up the process and drastically reduce the time and cost of drug discovery.

Project Goals:

- **Reduce Simulation Time**: Accelerate the event-based molecular dynamics simulations to deliver faster, real-time results for drug discovery teams.
- **Lower Costs**: Reducing the time required for simulations will reduce the overall cost of the drug discovery pipeline, particularly during the screening and testing phases.
- **Enable Scalability**: Allow larger simulations with more complex molecular interactions to be conducted efficiently using GPUs, improving the company's ability to test multiple compounds in parallel.

Technical Approach:

1. **Optimize Event-Based Molecular Dynamics (MD) Code**:
    - The current MD code is tailored to CPUs and lacks the parallelization capabilities required for modern drug discovery pipelines. Optimizing the code to exploit GPUs will unlock significantly higher levels of parallelism.
    - Implementation:
        i. Simulation is event-based: need to calculate/search for the timestep when the next collision occurs (N^2 time complexity). The two particles that collided have their data attributed to—the 1st chance for parallelism.

        ii. However, after the first collision, the other speeds are already calculated, so you only need to update the two particles that collided, and the other comparisons can be checked in (2N). 2nd chance for parallelism

2. **GPU Optimization with CUDA**:
    - **CUDA** will be used to implement low-level optimizations that enable fine-grained control over memory usage, thread management, and computation. This will ensure the code fully leverages the potential of NVIDIA GPUs.
    - Key optimizations will focus on reducing computational redundancy, improving memory throughput, and parallelizing the event-based structure of MD simulations.
3. **Testing & Validation**:
    - Performance tests will be conducted to compare the optimized GPU-based solution against the current CPU-based implementation in terms of speed, cost, and accuracy.
    - Validation will include a specific focus on the protein TDP-43 and how the GPU-optimized simulation performs in modeling its interactions with candidate drug compounds.
4. **Iterative Development and Integration**:
    - The rewritten code will be iteratively developed and integrated into Acelot's broader molecular simulation pipeline. Each iteration will involve feedback from Acelot's drug discovery team to ensure the tool meets their specific requirements.

Expected Benefits:

- **Time Savings**: Accelerating simulations could cut down weeks or even months from Acelot's drug discovery timeline, speeding up the design, synthesis, and screening of compounds.
- **Cost Efficiency**: Optimizing the MD simulations will lower operational costs and reduce the number of failed compound iterations.
- **Increased Throughput**: With GPU acceleration, Acelot can simulate more compounds in parallel, improving its chances of finding successful drug candidates faster.

Technologies:

- Python (NumPy, Matplotlib): mostly for testing and visualization
- CUDA (Math Libraries: CUDA-X GPU-Accelerated Libraries | NVIDIA Developer)
- AWS (GPU: A10, Linux: Ubuntu)

Timeline:

The project will follow a phased approach:

1. **Week 1:** Understand the problem deeper (run the repo and read the paper),
    a. Run a fewer number of collisions to speed up the process
    b. The paper is about the collision mathematics
    c. Let's finish the paper and prepare all the questions before our meeting on Friday
2. **Week 2:** Design Architecture Phase
    a. Create a backlog in the Spring
    b. Discuss what the first prototype will look like

Milestones

1. Treat atoms like billiard balls
    a. just calculate the velocity of each atom and the velocity after collision.
2. Add bonds for atoms
    a. The bond will act like a wall once the atoms go too far apart.
    c. Add the ability to create new bonds and break bonds
3. Implement Accuracy
    a. Pull statistics from the simulation runs and compare them to the current simulation implementation (single threading on CPU).

Process Model - Agile Methodology

- **Individuals and interactions** over processes and tools.
- **Working software** over comprehensive documentation.
- **Customer collaboration** over contract negotiation.
- **Responding to change** over following a plan.

Conclusion:

This project aims to significantly enhance the performance, accuracy, and cost-efficiency of Acelot's drug discovery efforts by rewriting its event-based molecular dynamics code for GPU optimization with CUDA. Ultimately, the project will help Acelot bring new neurodegenerative therapies to market faster and at a lower cost, benefiting patients and advancing research in this critical field.