A Distributed Approach for Reinforcement Learning Agents in High-Frequency Trading

Zimmer, R., Costa, O.

25 de Janeiro de 2025

Abstract

High-frequency trading (HFT) demands rapid decision-making and adaptability to dynamic market conditions. Reinforcement learning (RL) has emerged as a promising approach for developing data-driven trading strategies, but existing frameworks often struggle with scalability, latency, and daily updates of trading policies to adequately address changing market conditions. This thesis presents a novel distributed approach for an RL framework in the context of agent-based High-Frequency Trading (HFT). By introducing a multi-learner design and shared distributed limit order book (LOB) environments, the framework reduces redundant computations and individually simulated environments, achieving high scalability. Key performance metrics such as speedup, multicore efficiency, latency and loss and reward scores for trained agents are used to evaluate the framework's effectiveness for HFT applications. The proposed architecture aims to allow continuous training and deployment of adaptive agents, and therefore provide a practical approach towards more efficient and robust trading systems.

1 Introduction

1.1 Overview of High-Frequency Trading and Reinforcement Learning

High-frequency trading (HFT) represents one of the most technologically demanding domains within quantitative finance, characterized by its need for ultra-low latency, massive volumes of data, and the ability to adapt dynamically to rapidly changing market conditions. Reinforcement learning (RL) has emerged as a promising approach for developing trading strategies, offering the ability to learn optimal policies directly from market interactions without requiring explicit rules or models. However, applying RL to HFT is fraught with challenges, particularly when scaling training and inference to meet the performance requirements of live trading environments.

Traditional RL frameworks often rely on centralized architectures, where a single learner aggregates experiences from multiple actors to update a shared policy. While effective in general-purpose applications, such approaches can become bottlenecks in HFT contexts, where the scale and complexity of trading environments demand distributed efficiency and continuous adaptability. Moreover, typical RL systems treat environments as isolated entities for each agent or strategy, leading to significant computational redundancies when simulating complex environments like a limit order book (LOB). A LOB is a critical component in HFT systems, representing the collection of all outstanding buy and sell orders in a market, and its dynamics are crucial for understanding price formation and liquidity provision.

1.2 Limit Order Book Environments

A LOB is typically represented as a list of price levels, each containing the aggregated volume of orders at that price. For example, in a simplified LOB with four price levels, the bid side (buy orders) may look like this:

Price Level	Volume
101.71	30
101.84	50
101.95	50
102.01	10

Table 1: Bid Side of a Simplified Limit Order Book

Price Level	Volume
102.64	53
102.66	24
102.72	24
102.85	13

Table 2: Ask Side of a Simplified Limit Order Book

In this example, the bid side contains four price levels with corresponding volumes of 140, 110, 60, and 10, while the ask side contains four price levels with volumes of 53, 24, 24, and 13. The difference between the best bid and

ask prices is known as the bid-ask spread, which in this case is $\Delta = 102.01 - 102.64 = -0.63$. The LOB dynamics are continually changing as new orders arrive, are matched, or are canceled, reflecting the market's liquidity and

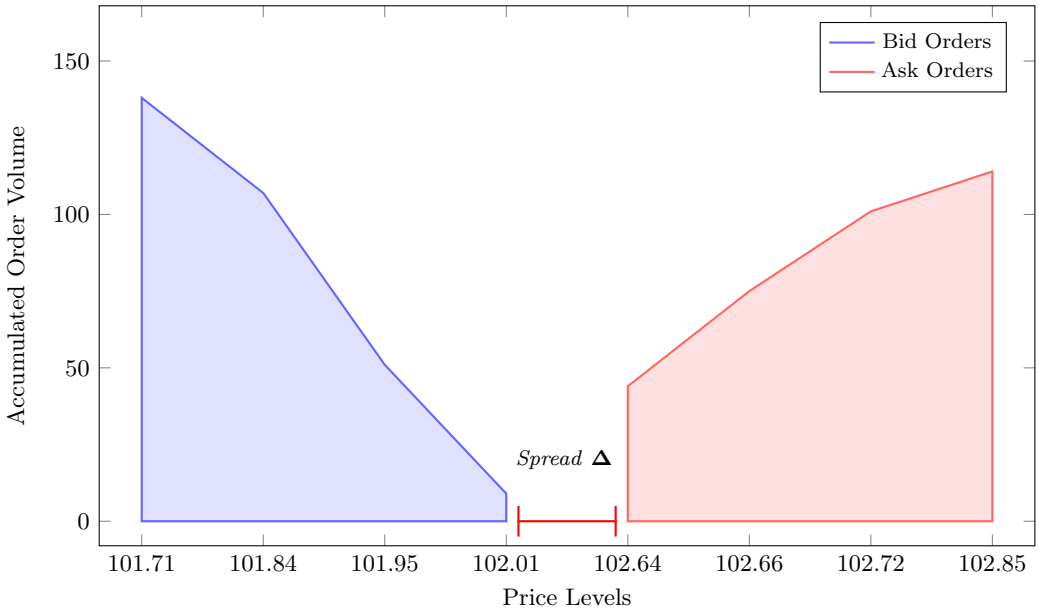

Figure 1: Limit Order Book Depth visualization for an arbitrary LOB $\mathbf{L}$

1.3 Motivation and Contributions

This thesis proposes a novel parallel, agent-based RL framework designed specifically for the HFT domain. The framework leverages Ray RLlib to parallelize environments and adopts a structure inspired by the actor-learner model introduced in the IMPALA (Importance Weighted Actor-Learner Architecture) algorithm. However, it extends this paradigm by allowing multiple learners to share parallelized environments, each corresponding to a unique agent definition, and enabling workers from different learners to share the same LOB environment. This design reduces computational overhead by reusing core environment components, such as order aggregation, order book initialization, and transaction processing, across multiple agents, and therefore reduces memory consumption and processing latency.

The primary contribution of the thesis is to provide a scalable, efficient, and adaptable framework for training RL agents in HFT applications, with the following key features:

- A parallel agent-based framework: A pragmatic implementation of a system architecture that enables efficient and decoupled training of multiple RL agents by leveraging distributed computing resources.
- Shared limit order book environments: An innovative mechanism for reducing redundant computations by allowing multiple workers, assigned to different learners, to interact with shared reusable LOB environment.
- Continuous train-and-deploy pipeline: The framework aims to support dynamic training and deployment of multiple strategies, where real-time updates to agent policies can be made without interrupting the trading process.

The framework must scale effectively with the number of agents, learners, and environments, maintaining consistent performance as workload increases. Performance metrics will include speedup and multicore efficiency, processing latency per action and end-to-end latency from state observation to order sending. The framework will be evaluated for its ability to support diverse trading strategies and market conditions through strategy-agnostic performance and modular design, enabling the reconfiguration of agent definitions and environments with minimal overhead. Additionally, the feasibility of deployment in live trading systems will be assessed through backtesting with simulated market data, where agents will be evaluated based on their loss and reward scores, and therefore assess their ability to adapt to changing market conditions.

The final thesis structure will first review relevant literature and novel approaches towards distributed agents in reinforcement learning and intersections with high-frequency trading, specifically parallel agent-based frameworks. An adequate search string will be fabricated, and a systematic review will be conducted to identify the most relevant literature, with a final discussion on the tagged literature and the relevance of each foundational paper to the proposed framework. A rigorous discussion into the problem formulation and system design, providing a detailed description of the proposed state, action and reward spaces will be made. The implementation of the framework will be discussed next, covering technical aspects and the integration of various components, such as the libraries used for message passing between workers and learners, policy networks and environment simulation dynamics. Experimental results will be presented to evaluate the performance and scalability of the framework, both in the computational and financial domains. Finally, the thesis concludes with a summary of results and a discussion of potential future directions for further research and development in the field.

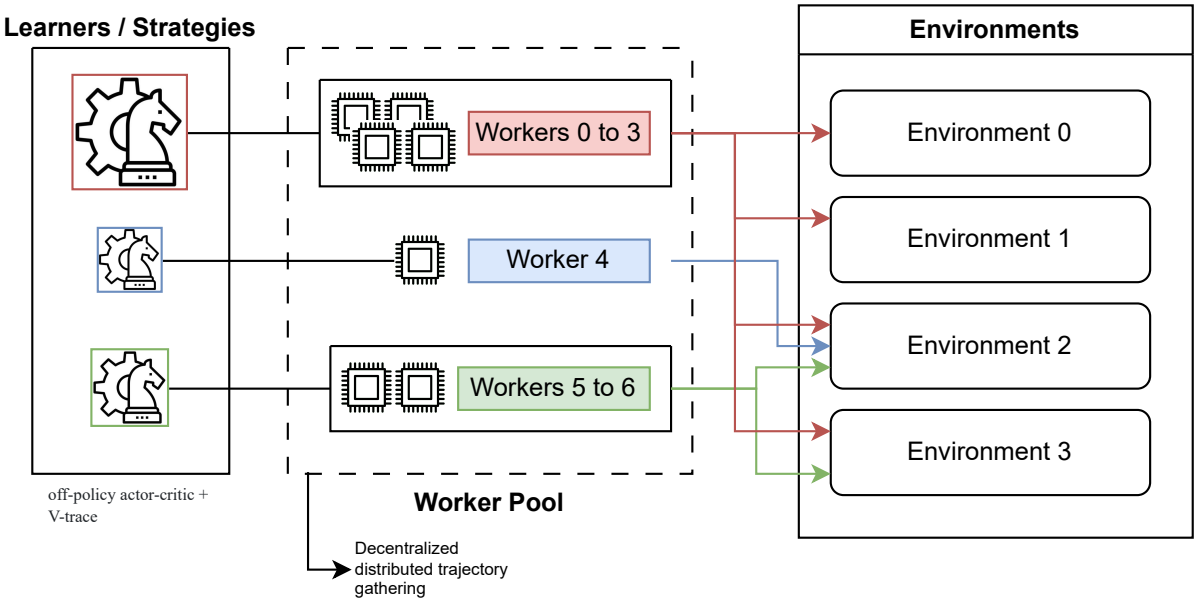

Figure 2: Diagram for the proposed parallel agent-based framework, with worker-learner interactions and shared limit order book environments.