Department of Electrical Engineering,
National Tsing Hua University

Special Topic on Implementation
Research Abstract

An FPGA-based Low Latency Limit Order Book Decoder and Its Visualizer

低延遲場效可程式化邏輯閘陣列之限價訂單簿解碼器及其可視化工具

Major Category: 系統領域

Group Number: B345

Advisor: Prof. Yeong-Luh Ueng

Members: Dang Hoang Khang Nguyen – Yang Ruei Chi

Research Period: From (2022/09/05) to (2023/11/13)

Abstract

In the realm of high-frequency trading (HFT) within financial markets, the demand for low-latency solutions to efficiently process real-time market data is crucial. This is especially crucial for understanding the dynamic nature of the financial market and making timely adjustments to trading strategies. Recognizing this importance, it becomes essential to reassess the approach to market data analysis. However, prevailing studies have predominantly focused on low-frequency datasets, leaving a significant gap in our comprehension of market dynamics within the high-frequency trading domain.

To address and bridge this gap, we present a pioneering solution featuring an FPGA-based decoder system with XGMII, UDP/IP, and TWSE data decoders, optimizing data throughput using Alveo U50 Data Center Accelerator Card. Operating at 312.5MHz, the FPGA-based XGMII decoder interfaces efficiently with the stock exchange's data feed, ensuring rapid data transfer. The UDP/IP decoder ensures reliable, low-latency data retrieval, and the TWSE data decoder handles the proprietary format. Verified on FPGA, our system accurately decodes TWSE financial market data, creating an order book with an aggregated latency of 0.26μs to 0.4136μs, a significant improvement over typical software-based system.

Innovatively, our system incorporates a limit order book visualizer, offering a real-time, intuitive representation of market dynamics. This visualizer enhances decision-making capabilities without the need for extensive background knowledge, providing traders with a valuable tool to optimize strategies in the fast-paced landscape of high-frequency trading.

摘要

在金融市場的高頻交易(HFT)領域中，以低延遲的方法去有效處理實時市場數據是極為關鍵的，它能使我們更好的理解金融市場的動態性或是及時調整交易策略。然而因現有研究多集中在低頻領域上，使我們不但在高頻交易市場動態的理解上有許多不足，現有的高頻交易市場分析方法也有不少需要調整之處。

為了解決低延遲的問題，我們在專題中實作了一個在 FPGA 上的解碼器系統，其中包含 XGMII 解碼器，UDP/IP 解碼器以及 TWSE 數據解碼器，並使用 Alveo U50數據中心加速卡來優化數據的吞吐量。XGMII 解碼器會和股票交易所的數據端進行有效連接，確保數據的快速傳輸，UDP/IP 解碼器能確保數據的可靠性和低延遲性，而 TWSE 數據解碼器則會對資料進行處理，將其中特定格式的資料輸出。經過在 FPGA 上的驗證，我們的系統能準確解碼 TWSE 的金融市場數據，且生成總延遲為0.26～0.4136微秒的訂單簿，與市場以往所採用的純軟體系統相比有顯著的進步。

此外，我們的系統融入了一個新的限價訂單簿視覺化工具，它能直觀的呈現實時市場動態。這個工具能在高頻交易快節奏的環境中增強交易者的決策能力，使交易者無需過多的背景知識，就能調整並優化其交易策略。

TABLE OF CONTENTS

List of Figures

List of Tables

1 Background

Nowadays, with the rapid development of science and technology, trading methods in the financial market have evolved from human-driven systems to computerized trading. High frequency trading (HFT) is a method of trading using computational power of computer programs to earn extremely short-term tiny price differences in the market at a speed higher than milliseconds. In other words, high-frequency trading provides extremely high liquidity in the market and collects tolls from providing liquidity. Based on a study conducted in 2009, about 73% of trading volume in the U.S equity markets is completed by high-frequency trading [1]. The explanation of high-frequency trading's popularity can be traced back to its remarkable profitability. This is exemplified by a noteworthy statistic indicating that thirty one HFT traders gained over $33 million in trading profits from a single E-mini S&P 500 futures contract within a month [2]. Nevertheless, only a selected few traders, whose requests are transmitted to the exchange with exceptional speed, manage to achieve such levels of profitability. Consequently, those traders with faster execution speeds outperform their counterparts with slower transaction speeds. The heightened competition in the field of high-frequency trading can be attributed to the scarcity of lucrative opportunities and the increasing number of participants in this domain. This escalating competition has, in turn, underscored the paramount importance of speed in trading. According to a study, every milliseconds reduction in latency can improve the arbitrage profitability by more than 100 millions dollars a year [3].

2 Motivation

At the heart of high-frequency trading's evolution is the Limit Order Book (LOB). Most modern financial markets, including the Taiwan Stock Exchange, operate on a double auction mechanism centered around the LOB. Traders submit limit and market orders, with the LOB serving as a central record of awaiting executions at different price levels on both buy and sell sides. Unlike the historical focus on time-series stock prices alone, the LOB provides a rich, multi-dimensional dataset. In Taiwan's rapidly advancing electronic trading landscape, especially on the Taiwan Stock Exchange, the LOB plays a crucial role in shaping strategies, particularly for high frequency traders with short time horizons [4]. Recognizing the limitations of software-based approaches in handling Limit Order Books for high-frequency trading, particularly in coping with the swift transfer times of high-speed Ethernet, the financial industry is turning its attention to application-specific FPGA-based accelerators. These FPGA systems offer notable advantages, including enhanced computational power, flexibility, re-programmability, and notably lower latency. While several FPGA-based systems have demonstrated significantly reduced latency for the Taiwan Futures Exchange [5] [6], attention on similar solutions for the Taiwan Stock Exchange has been relatively scarce. This research, therefore, not only seeks to bridge this gap by implementing an FPGA-based low-latency LOB decoder system tailored specifically for the Taiwan Stock Exchange but also recognizes the

need for a complementary visualizer. In the high frequency trading domain, where split-second decisions are paramount, a visual representation of the LOB dynamics becomes crucial for traders to swiftly comprehend and act upon market movements. Therefore, our research aims to provide a comprehensive solution, coupling a low latency decoder with an insightful visualizer, to empower traders with enhanced tools for rapid decision-making in the dynamic environment of the Taiwan Stock Exchange.

3 Research Purpose

Main purposes of our Special Topic are listed below:

• Customize and implement the network stack decoder: purpose-built to efficiently receive market feeds from the Taiwan Stock Exchange with a streamlined pipelined design on FPGA to markedly reduce the high latency associated with common software-based systems.

• Financial protocol decoder: developed exclusively for the Taiwan Stock Exchange's Format 6, simplifying the parsing process by focusing on essential order book data exclusively.

• Design FPGA-based and Python-based limit order book: efficiently manages the top five bid prices, top five ask prices, and their corresponding volumes for a single stock listed on the TWSE. This design minimizes latency in limit order book handling compared to traditional software-based approaches. Recognizing FPGA's limitations in visual output processing, a parallel architecture is introduced in Python to ensure a comprehensive solution.

• Limit Order Book Visualizer: introduces a straightforward Python-based visualizer chart. Users can conveniently choose the stock ID and available trading date, generating a visual representation of the Limit Order Book specific to their selection.

4 Research Method

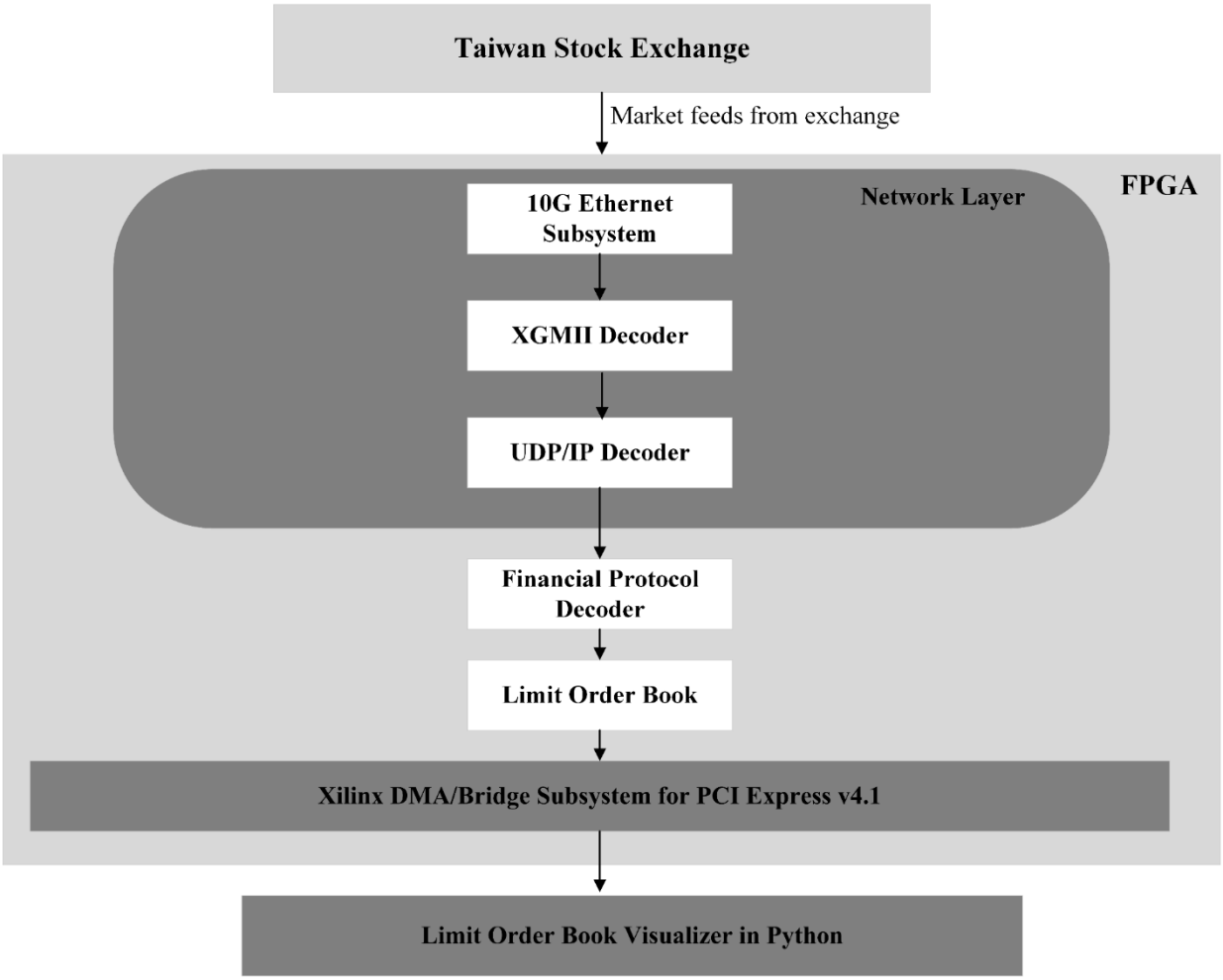

Fig 1. System architecture of Limit Orde 1

Our system comprises a network stack decoder, incorporating a 10G Ethernet subsystem, XDMII decoder, and UDP/IP decoder. The 10G Ethernet subsystem, driven by the 10G PCS-Only IP Core, enables high-speed reception and transmission of market data. The Ethernet decoder transforms packets for FPGA processing, and the UDP/IP decoder extracts headers. The financial protocol decoder parses TWSE Format 6, transforming data into a limit order book, including the top five bid/ask prices. The Xilinx DMA Subsystem acts as a seamless bridge between the FPGA and the host CPU for efficient data transfer. Finally, the visualizer of LOB is implemented using Python.

5 Experimental Result

Implementation evaluation is to check the offloading function of UDP by comparing the decoded messages using UDP/IP decoder on Vivado 2019.2 with Wireshark packets. Performance evaluation includes latency calculations ranging from approximately 0.26μs to 0.4136μs, visualized through dynamic animations of the Limit Order Book (LOB) on the x-axis for price levels and the y-axis for

trading volumes, providing a visualization of the LOB in one trading day.

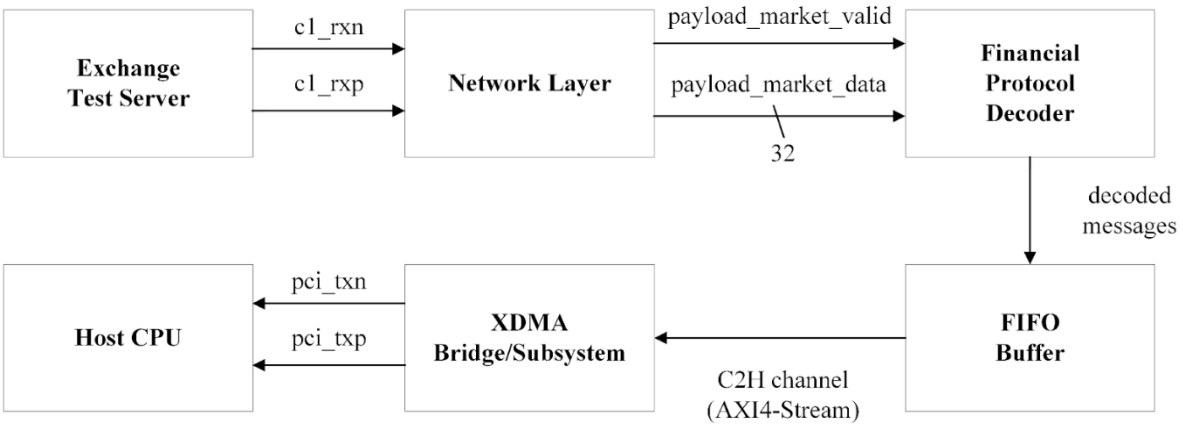

Fig 2. Block diagram of UDP/TCP offloading

Module	**Latency**	**Clock Cycle (Frequency)**
PCS-Only IP Core	0.0544 μs	17 (312.5MHz)
XGMII Decoder	0.0064 μs	2 (312.5 MHz)
UDP/IP Decoder	0.032 ~ 0.112 μs	10 ~ 35 (312.5 MHz)
TWSE Decoder	0.0288 ~ 0.1024 μs	9 ~ 32 (312.5 MHz)
XDMA Subsystem	0.124 μs	31 (250 MHz)
FIFO	0.0144 μs	
Overall	0.26μs ~ 0.4136μs	

Table 1 Latency inside each functional block of the system

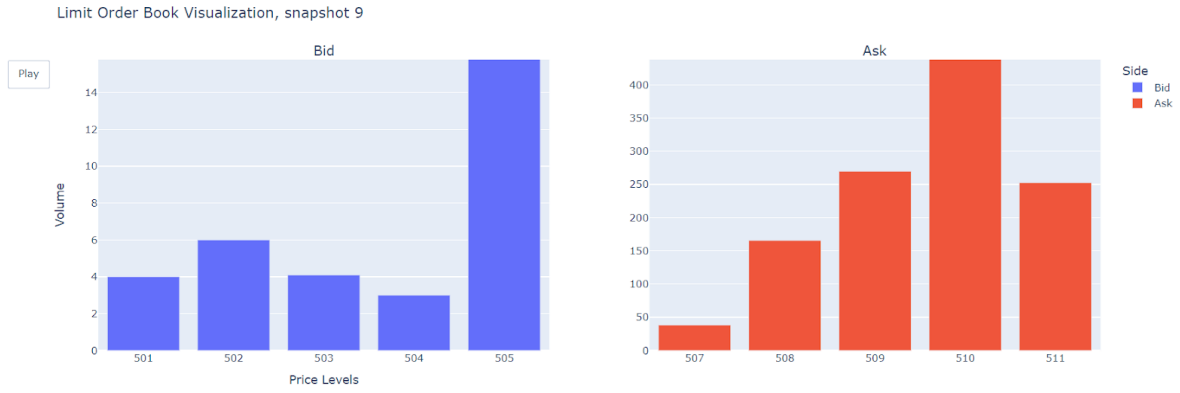

Fig 3. Limit Order Book Visualization from TWSE data on January 17th 2023 for TSMC stock (2330)

6 Conclusion

In this topic, an FPGA-based limit order book is implemented using Alveo U50 Data Center Accelerator Card on Vivado. The system provides the functionality of a network layer supporting UDP/IP protocol. Additionally, we implemented a financial protocol decoder tailored to Format 6's packet structure from the Taiwan Stock Exchange, allowing effective handling of the limit order book. To enhance user understanding, a straightforward animated visualization is presented using the Plotly library in Python. The overall system demonstrates notable efficiency, with a total end-to-end latency of approximately 0.26μs ~ 0.4136μs.

7 Reference

[1] Zhang, F. (2010). High-frequency trading, stock volatility, and price discovery. Available at SSRN 1691679.

[2] Baron, M., Brogaard, J., & Kirilenko, A. A. (2012). The trading profits of high frequency traders. Available at SSRN 2106158.

[3] Lockwood, J. W., Gupte, A., Mehta, N., Blott, M., English, T., & Vissers, K. (2012, August). A low-latency library in FPGA hardware for high-frequency trading (HFT). In 2012 IEEE 20th annual symposium on high-performance interconnects (pp. 9-16). IEEE.

[4] Bonart, J., & Gould, M. D. (2017). Latency and liquidity provision in a limit order book. Quantitative Finance, 17(10), 1601-1616.

[5] H. A Chen, “A Low Latency FPGA-Accelerated High-Frequency Trading System”, 2021 (student thesis)

[6] Y. C. Kao, “An FPGA-based High-frequency Trading System on Taiwan Futures Market”, 2021 (student thesis)

8 Review and Reflections

In the pursuit of developing an FPGA-based Low Latency Limit Order Book Decoder and its accompanying Python visualizer within the high-frequency trading (HFT) domain, this research journey has been an enriching and multifaceted experience. The amalgamation of knowledge we gained in the fields of HFT, FPGA programming, computer networks, Python, and Verilog has been instrumental in achieving the project's objectives. This endeavor allowed us to delve into the intricacies of HFT, honing skills in FPGA implementation for low-latency decoding, understanding the nuances of efficient computer network communication, and proficiently coding in both Python and Verilog.

We would like to express our sincere gratitude to our teammates, whose collaborative efforts played a pivotal role in the success of this project. The invaluable guidance from the master students from the ECC Lab, fostered an environment of shared knowledge and expertise. We are especially thankful for the guidance and mentorship provided by our supervisor, Prof. Yeong-Luh Ueng, whose insightful feedback and unwavering support significantly influenced the project's trajectory. Their collective commitment not only enriched our understanding of the subject matter but also cultivated a sense of teamwork that transcends the technical aspects of the research. This experience has not only expanded our technical proficiency but also underscored the importance of collaborative endeavors and mentorship in the realm of academic research.