

2025-2026 Master Thesis Topics in OpenMP and ML/AI

John Burns, ADA University

July 15, 2025

Overview

- 1 Introduction
- 2 Topic 1: Optimizing Deep Learning Inference with OpenMP
- 3 Topic 2: Parallelizing Reinforcement Learning Algorithms with OpenMP
- 4 Topic 3: OpenMP for Energy-Efficient ML Training on Multi-Core Systems
- 5 Topic 4: Hybrid OpenMP-MPI Frameworks for Distributed ML Workloads
- 6 Topic 5: OpenMP for Real-Time ML Inference in Edge Devices

About myself and my research interests

- My own PhD was in the field of large-scale (at the time!) modelling and simulation
- From my industry background I also have an interest in performance, reliability and scalability
- From my teaching I have an interest in Computer Architecture and Distributed Systems
- I welcome any suggestions you may have in the intersection of these fields
- I am very interested in OpenMP projects - from code and performance analysis through to application developement
- As Energy Efficiency and Optimization is so important now, why not use ARM64 cloud platforms for these research topics.

Topic 1: Optimizing Deep Learning Inference with OpenMP

- **Objective:** Enhance the performance of deep learning inference on multi-core CPUs using OpenMP parallelization techniques.
- **Description:** Investigate how OpenMP can be used to parallelize neural network inference tasks, focusing on convolutional neural networks (CNNs) and recurrent neural networks (RNNs). Explore loop parallelization, task scheduling, and data partitioning to reduce inference latency.
- **Research Questions:**
 - How can OpenMP directives optimize matrix operations in deep learning models?
 - What are the trade-offs between OpenMP task-based and loop-based parallelization for inference?
- **Expected Outcomes:** A framework for OpenMP-based optimization of deep learning inference, with performance benchmarks on standard datasets like ImageNet.

Topic 2: Parallelizing Reinforcement Learning Algorithms with OpenMP

- **Objective:** Improve the training efficiency of reinforcement learning (RL) algorithms by leveraging OpenMP for parallel computation.
- **Description:** Focus on parallelizing key components of RL algorithms, such as Q-learning or policy gradient methods, using OpenMP. Study the impact of parallel environment simulations and gradient computations on training speed.
- **Research Questions:**
 - How can OpenMP parallelize environment interactions in RL frameworks like OpenAI Gym?
 - What are the scalability limits of OpenMP in distributed RL settings?
- **Expected Outcomes:** A scalable OpenMP-based RL training pipeline with performance comparisons against serial implementations.

Topic 3: OpenMP for Energy-Efficient ML Training on Multi-Core Systems

- **Objective:** Develop energy-efficient machine learning training workflows using OpenMP on multi-core architectures.
- **Description:** Explore OpenMP's thread management and scheduling capabilities to minimize energy consumption during ML model training, focusing on algorithms like gradient descent for large-scale datasets.
- **Research Questions:**
 - How can OpenMP dynamic scheduling reduce energy usage in ML training?
 - What is the impact of thread affinity on energy efficiency in multi-core CPUs?
- **Expected Outcomes:** A set of OpenMP-based techniques for energy-efficient ML training, validated with power consumption metrics.

Topic 4: Hybrid OpenMP-MPI Frameworks for Distributed ML Workloads

- **Objective:** Design a hybrid OpenMP-MPI framework to optimize distributed machine learning workloads across clusters.
- **Description:** Combine OpenMP for intra-node parallelization with MPI for inter-node communication to enhance the scalability of ML algorithms like distributed stochastic gradient descent (SGD).
- **Research Questions:**
 - How can OpenMP and MPI be integrated to balance computation and communication overheads?
 - What are the performance benefits of hybrid parallelization in distributed ML?
- **Expected Outcomes:** A hybrid OpenMP-MPI framework for distributed ML, with case studies on large-scale datasets.

Topic 5: OpenMP for Real-Time ML Inference in Edge Devices

- **Objective:** Enable real-time machine learning inference on edge devices using OpenMP parallelization.
- **Description:** Investigate the use of OpenMP to optimize ML inference on resource-constrained edge devices, focusing on lightweight models like MobileNet. Study task partitioning and thread management for low-latency inference.
- **Research Questions:**
 - How can OpenMP optimize inference on edge devices with limited cores?
 - What are the performance trade-offs of OpenMP in real-time ML applications?
- **Expected Outcomes:** An OpenMP-based inference framework for edge devices, with latency and throughput benchmarks.

Conclusion

- These topics combine OpenMP's parallel computing capabilities with cutting-edge ML/AI challenges.
- Each topic offers opportunities for performance optimization, scalability, and real-world impact.
- Recommended tools: OpenMP 5.0, TensorFlow/PyTorch, and multi-core CPU clusters.
- OpenMP Spec OpenMP documentation for further details.