Matthew Ernst

| 970-214-5508 | matthewernst@apple.com | matternst.dev |
| linkedin.com/in/matthew-f-ernst | github.com/matthewfernst |

Experience

Software Engineer - Foundation Models Framework October 2025 - Present
Apple Inc *Cupertino, CA*

Software Engineer - Sensing and Connectivity October 2024 - October 2025
Apple Inc *Cupertino, CA*

Senior Software Engineer, Machine Learning May 2022 - October 2024
Qualcomm, Machine Learning Group - AIMET/AISW Core Tools Team *San Diego, CA*

- Lead optimizations for LLM/LVMs such as LLaMA and Stable Diffusion by contributing to the development of AIMET, an open-source library focused on implementing advanced quantization and compression techniques for trained neural network models. AIMET reduces latency while maintaining original accuracy within 1% when running models on AI dedicated Qualcomm hardware.
- Created PyTorch Model Preparer Pro, reducing overall AIMET development time by up to 4 hours per model iteration. PMMP takes hardware optimized models and reconstructs them in AIMET, thus mitigating the need to reconvert models for hardware every iteration.
- Implemented Keras Per Channel Quantization and Quantization Aware Training (QAT) support within AIMET, directly unblocking the image computation pipeline for flagship phones including the Samsung Galaxy S23 and beyond.
- Supported Microsoft in efforts with Windows on Snapdragon by reimplementing Batch Normalization folding and re-estimation of models, fixing AutoQuant training time decisions for correct quantized models, and redesigning Tensor Quantizer for better scalability to newer quantization techniques.

Projects

Lynx (iOS) / Mountain UI (Electron) December 2022 – Present
- Developed a native iOS app using Swift, integrating with the Slopes app, to upload devices to a custom AWS Lambda API. Utilized GraphQL client to showcase user statistics and leaderboard information amongst friends.
- Built an Electron app that enhances the skiing experience by providing real-time updates of lifts and trails at ski resorts, including weather forecasts, live cams, and dynamic leaderboard based on Lynx.

Chord - A Peer to Peer System September 2021 – December 2021
- Successfully created and implemented a distributed file system in Python, leveraging the Chord protocol to ensure equal workloads and efficient data partitioning across the network.
- Designed and implemented a hashable 16-bit ID space, enabling the accurate storage and retrieval of up to 64,000 peers and keys within the Chord ring, ensuring scalability and optimal performance.

Education

Master of Science in Computer Science December 2021
Colorado State University, GPA: 4.0 *Fort Collins, CO*
Research: Sparse Reconfigurable Artificial Neural Systems *May 2021 – May 2022*
- Researched the underlying structure of ReLU networks and the presence of dead neurons from vanishing gradients.
- Implemented new neural network architecture to mitigate dead neurons named a "Late Residual Neural Network."
- Investigated correlations between learning rates and optimizers to an increased quantity of dead neurons.

Bachelor of Science in Biological Systems Engineering December 2018
Iowa State University *Ames, IA*

Technical Qualifications

Languages: Python, C++, Swift, C, Objective-C, Java, JavaScript, Rust
Frameworks / Libraries: PyTorch, Keras, TensorFlow(1/2), ONNX, JAX, MLX (Contributed), PyBind, UIKit, SwiftUI, AppKit, GraphQL, OpenCV, Node, React, Electron, Apollo, Vite, Webpack, Jest, JUnit, Maven, Gradle, MongoDB
Tools: Linux, Git, Vim, Scrum, Docker, Postman, Jenkins, GCP, AWS, JetBrains, Visual Studio, Serverless
Engineering Principles: Agile Development, Object Oriented Programming, Cloud Computing, Test Driven Development, Unit Testing, Coverage Testing, Continuous Integration/Deployment