

This dissertation presents an in-depth analysis of using heterogeneous computing to improve the performance of AI algorithms and scientific simulations. This study focuses on two key project approaches: 1) the parallelization of high-compute large matrix multiplications with complex values, including large Hermitian matrix multiplications and complex SVD (Singular Value Decomposition) calculations, with the results of SVD being used in machine learning tasks, and 2) the parallelization of Shapley value computation for Explainable AI (XAI) applications. These approaches are thoroughly benchmarked across multiple hardware configurations, not limited to consumer-grade hardware but also including high-performance data center GPUs and CPUs. The first study focuses on parallelization methods within CUDA and DPC++ SYCL programs. This study also illustrates the benefits of transitioning from vendor-specific programming models to more universal standards like C++ SYCL. These parallelization techniques were implemented on diverse hardware platforms and comprehensively benchmarked to evaluate the impact of parallel programming on optimizing performance. The second study applies Shapley values in the field of XAI to deliver crucial understanding into the decision-making processes of deep learning models trained on image datasets. This research highlights how connecting advanced hardware architectures with sophisticated algorithms can revolutionize heterogeneous computing to handle some of the most demanding tasks in modern AI and scientific research.