AI Acceleration using Graphcore IPUs

Machine learning and neural networks are used extensively for all kinds of research nowadays. During the last decade, they have established themselves as the mainstream method in processing vast amounts of data, while reusing similar models/architectures in entirely different fields. In order to reach this point, major breakthroughs were required on the computing architectures used to train large networks efficiently. The most popular processing units used today to train and infer AI models are the GPUs.

Since then, more competitors have entered the market offering hardware dedicated explicitly in the acceleration of AI applications. In a recent advent, Graphcore introduced its Intelligence Processing Units (IPUs), specialized in executing the heavy workload of Machine and Deep Learning applications. The IPUs are based on a different memory and compute hierarchy compared to GPUs, utilizing a distributed memory model along their thousands of cores, which run in parallel. Each IPU can be connected to more IPUs using high-speed links, ultimately forming a scalable network of compute units.

In this work, you will focus on exploring the capabilities of Graphcore IPUs on an AI project. Not only will you get hands-on experience with Machine Learning, but you will also gain insights on how cutting-edge computer architectures handle the compute and memory limitations of neural networks today. The goal is to adopt an existing ML application's code in high-level libraries (Tensorflow, Pytorch) which works on GPUs and reproduce it to work on IPUs. Then, extensive profiling and comparison between the two platforms will follow. Ideally, the candidate will also experiment with Graphcore's graph-programming framework, Poplar SDK, which has a C++ interface.

PREREQUISITES:

Strong knowledge of **Bash**, **Python**.

Desirable: Familiarity with **Machine Learning, Deep Learning, C++.**

SKILLS YOU WILL LEARN:

Hands-on experience on a deep-learning project, tensorflow and/or pytorch, familiarization with graphcore hardware and tools.

RELATED MATERIAL:

5 Reasons why we need new machine learning hardware
(https://www.graphcore.ai/posts/5-reasons-why-we-need-new-machine-learning-hardware)

IPU Programmer's Guide (https://docs.graphcore.ai/projects/ipu-overview/en/latest/index.html)

CONTACT INFORMATION:

Petros Arvanitis, Researcher, NTUA (parvanitis@microlab.ntua.gr)

Associate Professor Christos Strydis, Erasmus Medical Center & Delft University of Technology (c.strydis@erasmusmc.nl)

Professor Dimitrios Soudris, NTUA (dsoudris@microlab.ntua.gr)