LambdaGen - A GPU code generator powered by recursion schemes

András Leitereg

Faculty of Informatics, Eötvös University, Budapest

Problems that are data and computation intensive but are also naturally parallelizable are very frequent in physics, statistics and other fields of science. Their efficient parallel implementation would however require not just decent programming skills, but also deep insight of the target architecture. These computations are often in the form of a linear algebraic expression, and our goal is to deduce their optimal parallelization and memory usage automatically. We'd like to create a tool that offers a compromise between minimizing the implementation and the running time, for scientists who are not programmers.

Most of the existing linear algebra libraries are hardly usable for complex high dimensional computations. Partly because the offered operation primitives (e.g. matrix multiplication) can't be modified or parameterized, and the data representation is often not general either (e.g. they don't support structures over 3 dimensions). Our work includes an alternative representation of the linear algebraic expressions, which lifts these restrictions and allows us analyze and optimize the expression tree itself.

I'll present the LambdaGen tool, which is the continuously developed implementation of our research. It applies a series of recursion scheme based transformations on an expression tree, and produces SYCL code to evaluate the expression on GPU.