

doi:10.1631/FITEE.1500032

题目: 使用“基于分析的代码转换方法”来提升 GPU 特定的 OpenCL kernel 在多核/众核 CPU 上的性能移植性

目的: 针对面向 GPU 设计的 OpenCL kernel 程序在 CPU 上性能移植性欠佳这一问题, 设计一种基于访存特征分析的代码转换方法, 提升性能移植性。

创新点: 通过分析 OpenCL kernel 中的访存模式, 去除不必要的局部存储数组及其带来的同步语句, 并使用向量化和局域性重开发进一步优化代码, 最终取得显著的性能提升。

方法: 首先, 针对 OpenCL kernel 代码中的数组访问, 设计一种精确的线性化访问描述子(图 2)。然后, 利用该描述子, 分两步对 GPU 特定的 OpenCL kernel 代码进行转换, 以提高其在 CPU 上的性能(图 7)。第一步为基于分析的 work-item 折叠, 即通过分析访问描述子, 找出并去除不必要的局部存储数组及其带来的同步语句, 然后完成 work-item 折叠。第二步为适应架构的代码优化, 即针对 CPU 架构的特点, 使用向量化和局域性重开发进一步优化折叠后的代码。最后, 上述代码转换过程被整合为一个工具链, 连同调度程序, 嵌入到一个开源的 OpenCL 运行时系统中(图 11)。实验结果表明, 这种转换方法可以显著提升 GPU 特定的 OpenCL kernel 在 Intel Sandy Bridge 架构 CPU 和 Intel Knights Corner 架构协处理器上的性能。

结论: 准确分析 OpenCL kernel 代码中的访存模式, 不仅利于判断局部存储数组是否适合于 CPU 架构, 还能用于指导之后的代码优化过程, 因此是提高性能移植性的重要步骤。

关键词: OpenCL; 性能移植性; 多核/众核 CPU; 基于分析的转换