

Полиэдральный кодизайн тайлинга циклов в MLIR

А.В. Левченко

Суперкомпьютерный центр СПбПУ

Актуальность работы обусловлена необходимостью разработки промежуточного языка планирования для тайлинга циклов. Тайлинг — это метод разделения вычислительных операций алгоритма на подмножества (тайлы) для реорганизации порядка вычислений, применяемый в целях оптимизации локальности и параллелизма программ в области HPC.

Научная новизна. В работе предложен метод совместного проектирования тайлинга и слияния циклов, называемый полиэдральным кодизайном (рис. 1). Метод решает проблему согласованности тайлинга и слияния циклов в рамках полиэдральной компиляции [1]. *Входными данными* метода являются циклы вычислительной программы на языке многоуровневого промежуточного представления (Multi-Level Intermediate Representation, MLIR) библиотеки LLVM, а также размеры тайлов, полученные посредством алгоритма Tile Size Selection (TSS). Совместное проектирование тайлинга и слияния циклов выполняется посредством полиэдральной компиляции с использованием диалекта MLIR/Affine. Далее реализуется описание трансформирующих операций на языке планирования (Tiling Intermediate Representation, TIR) на основе диалекта MLIR/Transform. Язык TIR применяется при метапрограммировании диалектов для тайлинга посредством диалекта MLIR/IRDLL. Наконец, полученные варианты трансформаций применяются к IR целевой программы. *Выходными данными* метода являются трансформированные циклы на языке MLIR.

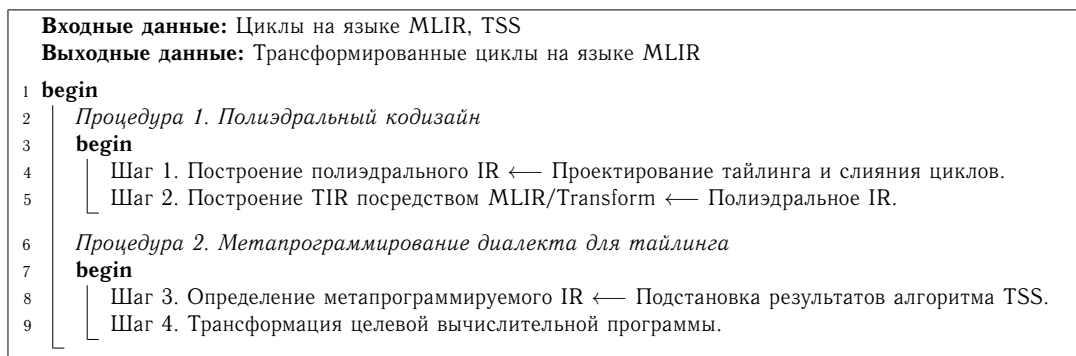


Рис. 1. Процедура полиэдрального кодизайна в конвейере тайлинга циклов MLIR

Экспериментальные результаты получены для полиэдральных версий тестовых программ класса трафаретных вычислений (stencils) FDTD-1D/2D, Jacobi-1D/2D и Seidel-2D при выполнении на системе с AMD EPYC 7763. Определен диалект для тайлинга циклов в IR используемых тестовых программ, позволяющий включить результаты алгоритма TSS в качестве параметров размеров тайлов для операций Transform IR. В результате полиэдральный кодизайн тайлинга и слияния циклов, реализованный посредством языка планирования MLIR/Transform с использованием значений TSS, позволил получить более чем двукратное ускорение тайлового кода относительно функции планирования для базового тайлинга.

Литература

1. Zhao J., Xu J., Di P. et al. Modeling the Interplay between Loop Tiling and Fusion in Optimizing Compilers Using Affine Relations // ACM Transactions on Computer Systems. January, 2024. Vol. 41, no. 1–4. Article 5. DOI: 10.1145/3635305