

令和元年6月22日現在

機関番号：32660

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K00107

研究課題名(和文) COINS共通基盤のPTX向け拡張

研究課題名(英文) Extension of COINS Compiler Infrastructure Based on PTX

研究代表者

滝本 宗宏 (Takimoto, Munehiro)

東京理科大学・理工学部情報科学科・教授

研究者番号：00318205

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：本研究では、コンパイラ共通基盤COINSの中間表現LIRから、GPU用コードPTXを生成することによって、COINS上でGPU向けの低水準最適化を開発できる環境を実現した。その過程で、LIRの配列を含む主要計算部をPTXへ変換する変換器をCOINSコンパイラ共通基盤に実現し、GPUで動作することを確認した。さらに、分岐発散を低減する手法を4つ実現し、最終的に、最大1.8倍の実行効率向上を実現した。

研究成果の学術的意義や社会的意義

GPU特有の問題に基づいたコード最適化器を、COINSに既存のLIR上で開発できるようにすることは、複雑化しているGPU向けコード最適化器の開発を容易にする効果がある。特に、近年注目され始めた分岐発散の問題に関する進展は、大規模並列計算の性能向上に大きく貢献するものと考えられる。また、従来各中間表現に固定だったコード最適化器が、LLVM-IRとPTXの両方に適用できるようになることで、GPU向けコード最適化に、新しいチューニング法の道を開く。

研究成果の概要(英文)：In this research, we aimed at making implementation of GPU optimization easier by extending the COINS compiler infrastructure by translation from LIR (Low-level Intermediate Representation in COINS) to PTX. In this process, we implemented a translator from LIR to PTX, proving that the generated PTX works well on a GPU. Also, we proposed four approaches for suppressing branch divergence, achieving the maximal gain of 80%.

研究分野：言語処理系

キーワード：分岐発散 GPU 部分冗長除去 投機的移動 疎なコード移動 PTX

## 様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

### 1. 研究開始当初の背景

グラフィックス処理を高速化するために開発された GPU(Graphics Processing Unit) は、その大量のデータを複数のプロセッサで並列実行する性質から、近年、高性能計算で多用されるようになってきた。一方で、GPU が実行性能を発揮するためには、コストが高いメモリ参照の低減や、Warp と呼ばれるスレッドグループが、一致した条件分岐経路を実行しない場合に、分岐先を逐次実行する分岐発散 (branch divergence) の緩和が重要である。

当該研究者は、今年度までの基盤研究(C) において、コンパイラ共通基盤 COINS を、GPU 向け言語処理系の多くが採用している中間表現 LLVM-IR 上の変換器として拡張するとともに、GPU および CPU におけるメモリ参照を、大域的なコード移動を用いて、キャッシュやレジスタ参照に置き換えるコード最適化の実現を行ってきた。一方で、分岐発散を緩和する手法を検討する過程で、LLVM-IR 上で実装することの難しさが明らかになってきた。

本研究では、より GPU のコードに近い中間表現である PTX 向けに、COINS を拡張し、分岐発散に関連する、より効果的なコード最適化を実現するとともに、それらの効果を、最適化器に変更を加えることなく、LLVM-IR 上および PTX 上の両方で、評価できるようにする。

### 2. 研究の目的

当該研究者は、GPU 向けプログラミング言語 CUDA や Open-CL の中間表現としても採用されている LLVM-IR 上のコード最適化開発を容易にするために、コンパイラ共通基盤 COINS を LLVM-IR 向け変換フィルタとして拡張してきた。COINS は、LLVM-IR より単純な中間表現 LIR を備えているので、GPU 向けコード最適化器を容易に実現できるとともに、LIR 向けの多くのコード最適化を利用することができる。しかしながら、LLVM-IR 上の変換では、GPU 特有の性質を活かしたコード最適化を実現できないことが分かってきた。本研究では、COINS を、より低水準な中間表現 PTX についても拡張し、LIR 上の最適化器を、LLVM-IR と PTX の両方で評価可能にする。

### 3. 研究の方法

初年度は、分岐発散を低減する手法を中心に、GPU 特有の命令を用いる手法について文献を調査し、同時に、COINS の LIR において、 $\phi$  関数を任意の場所に挿入できるように拡張する。中間年度は、COINS の入力および出力として PTX を利用できるように拡張し、さらに、分岐発散を低減する手法を、拡張した COINS 上に実現する。最終年度では、実現したコード最適化を、PTX と LLVM-IR の両方で性能評価するとともに、両最適化器の共通化を検討する。

### 4. 研究成果

本研究では、コンパイラ共通基盤 COINS の中間表現 LIR から、GPU 用コード PTX を生成することによって、COINS 上で GPU 向けの低水準最適化を開発できる環境を実現した。その過程で得られた研究成果は次のとおりである。

#### (1) 大域値番号付けを用いたくり出しに基づく分岐発散低減法の実現

一度プログラムを走査するだけで多くの冗長な式を除去できる大域値番号付けという最適化手法に、if 文内に存在する式を、if 文の上方に移動する拡張を加えた手法である。大域値番号付けは、従来、if 文内に存在する共通部分式を除去することによって、分岐発散を低減できる。本手法では、さらに、if 文内の式を外に出すことによって、共通部分式を増加させる効果をもつので、より多くの分岐発散を低減できる。本手法は、if 文の then あるいは else 部のいずれかだけに存在する式でも、if 文の上方に移動する可能性がある。このような場合、逐次実行では、式が存在しなかった側にも計算コストを導入することになるので、実行コストが増加する。一方、GPU 上では、if 文の then 部も else 部も 1 度ずつ実行されるので、たとえ片方からしか式が移動しなかったとしても、実行コストが増加しない点が重要である。

#### (2) LIR の主要部分を PTX へ変換する変換器の実現

LIR の配列を含む主要計算部を PTX へ変換する変換器を COINS コンパイラ共通基盤に実現し、GPU で動作することを確認した。

#### (3) プロファイリングに基づく部分冗長除去を用いた投機的移動と分岐発散低減の実現

プロファイリングに基づく部分冗長除去を、分岐発散している条件分岐に対して故意に偏ったプロファイルデータを設定することによって、選択的な投機的移動を実現する。本手法を適用することによって、if 文の then 部も else 部も 1 度ずつ実行される分岐発散部に限って、if 文の上に命令を巻き上げるので、実行コストを増加させることなくプログラム全体の冗長性を増加させることができる。増加した冗長性は、命令を挿入しながら多くの冗長な式を取り除く部分冗長除去の効果によって除去するが、プロファイルデータの偏った設定によって、分岐発散を増加させる挿入を生じることなく、冗長性を除去できる。

#### (4) 疎なコード移動 (Sparse Code Motion, 以下 SCM と呼ぶ) と投機的移動を組み合わせた分岐発散低減手法の実現

部分冗長除去法の一つで、静的な式の数減らすように式を減らす疎なコード移動が分岐後の

式を分岐前に移動する効果と、分岐発散した分岐後のコードはすべて順番に実行される性質に着目し、疎なコード移動を投機的移動と組み合わせることによって、分岐内から分岐前に式を移動する効果を高めた手法 (Speculative SCM, 以下 SSCM と呼ぶ) を提案した。

**(5) プログラム中の分岐発散を生じている分岐と生じていない分岐に対して、分岐発散低減手法と部分冗長除去を選択的に適用する手法の実現**

分岐発散していない分岐における投機的移動が、プログラムの実行効率を低減する可能性があることに着目し、分岐発散している分岐としていない分岐を、分岐発散解析によって解析し、分岐発散している分岐には SSCM として、分岐発散していない分岐には SCM として、選択的に適用できる手法を提案した。SSCM の選択的な適用について、一般的に使われるサンプルコードで評価を行ったところ、1.8 倍の効率化を実現した。本成果は、第 21 回プログラミングおよびプログラミング言語ワークショップ (PPL2019) で発表済みである。現在、国際会議プロシーディングスおよびトランザクションに投稿すべく準備を進めている。

5. 主な発表論文等

〔雑誌論文〕(計 件)

〔学会発表〕(計 1 件)

福原 淳司, 滝本 宗宏, コード移動に基づく分岐発散低減, 第 21 回プログラミングおよびプログラミング言語ワークショップ (PPL2019)

〔図書〕(計 件)

〔産業財産権〕

出願状況 (計 件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
出願年：  
国内外の別：

取得状況 (計 件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
取得年：  
国内外の別：

〔その他〕

ホームページ等

6. 研究組織

(1) 研究分担者

研究分担者氏名：

ローマ字氏名：

所属研究機関名：

部局名：

職名：

研究者番号 (8 桁)：

(2)研究協力者

研究協力者氏名：

ローマ字氏名：

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。