

LLMはLLVM IRのsuperoptimizerになれるのか？

浅野 光平, 高前田 伸也 (東京大学 大学院情報理工学系研究科)

目的：practical なSIMD-oriented superoptimizerを実現したい

背景

- 近年のプロセッサのSIMD命令の進化を存分に生かしてプログラムの最適化を行いたい
 - SIMD命令はCPU毎のコストモデル, 多様なIntrinsicによる管理のコストが高い
- コンパイラのpeephole optimizationは有効だが保守管理が難しい

Superoptimizer

- 合成, 列挙的なアプローチで命令の変換を探索
- SMTソルバーで変換の正当性を検証し, 正しいことを証明

Souper: A Synthesizing Superoptimizer, [Raimondas Sasnauskas](#) et al.

Minotaur: A SIMD-Oriented Synthesizing Superoptimizer, [Zhengyang Liu](#) et al.

提案：LLMベースの反例駆動型合成

モチベーション

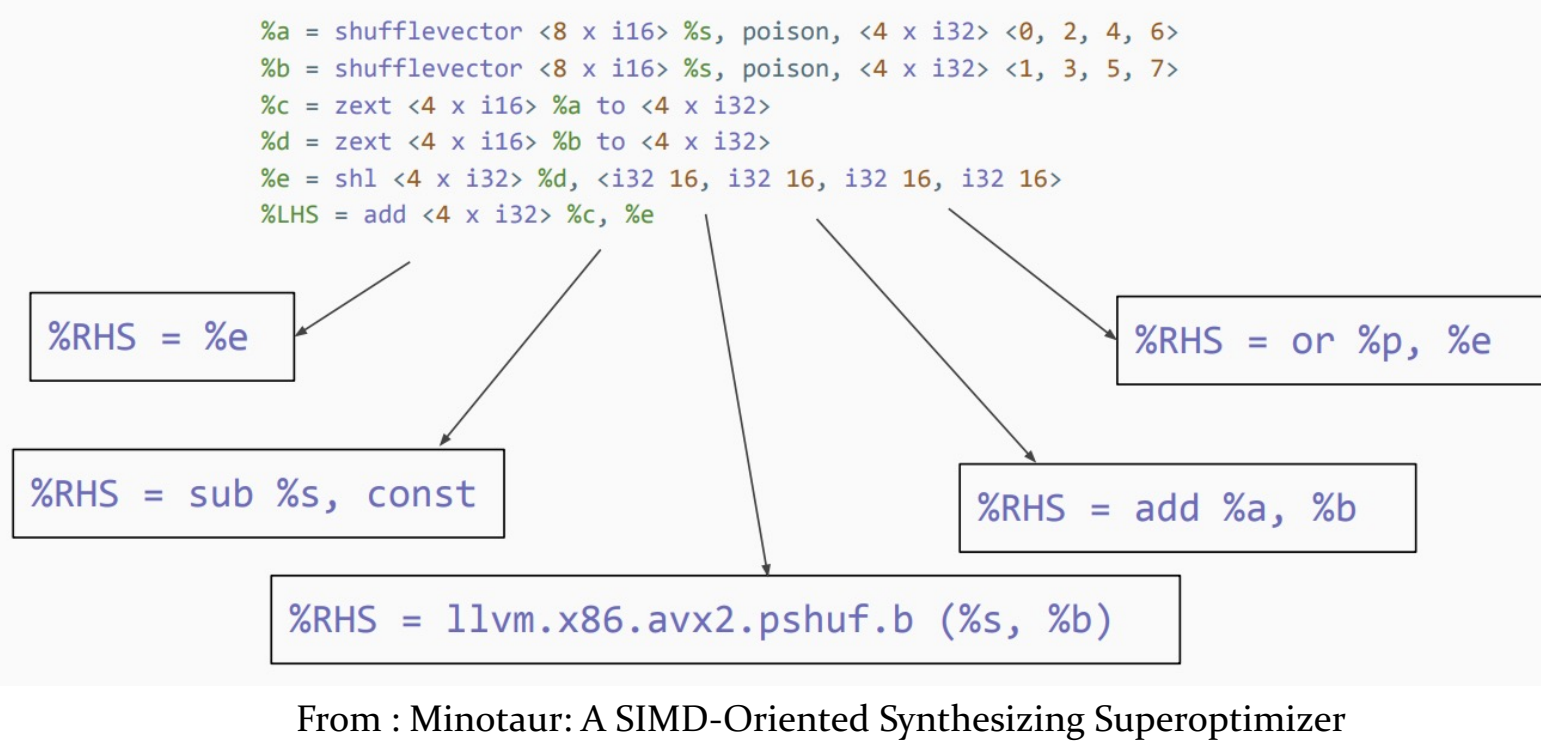
- Peephole最適化ならば個別の最適化Web上からのパターンマッチでそれらしい変換が可能ではないか
- 論理的なFeedbackがあれば, 言語モデルの論理的な問題に対する弱さを補えるのではないか

アプローチ

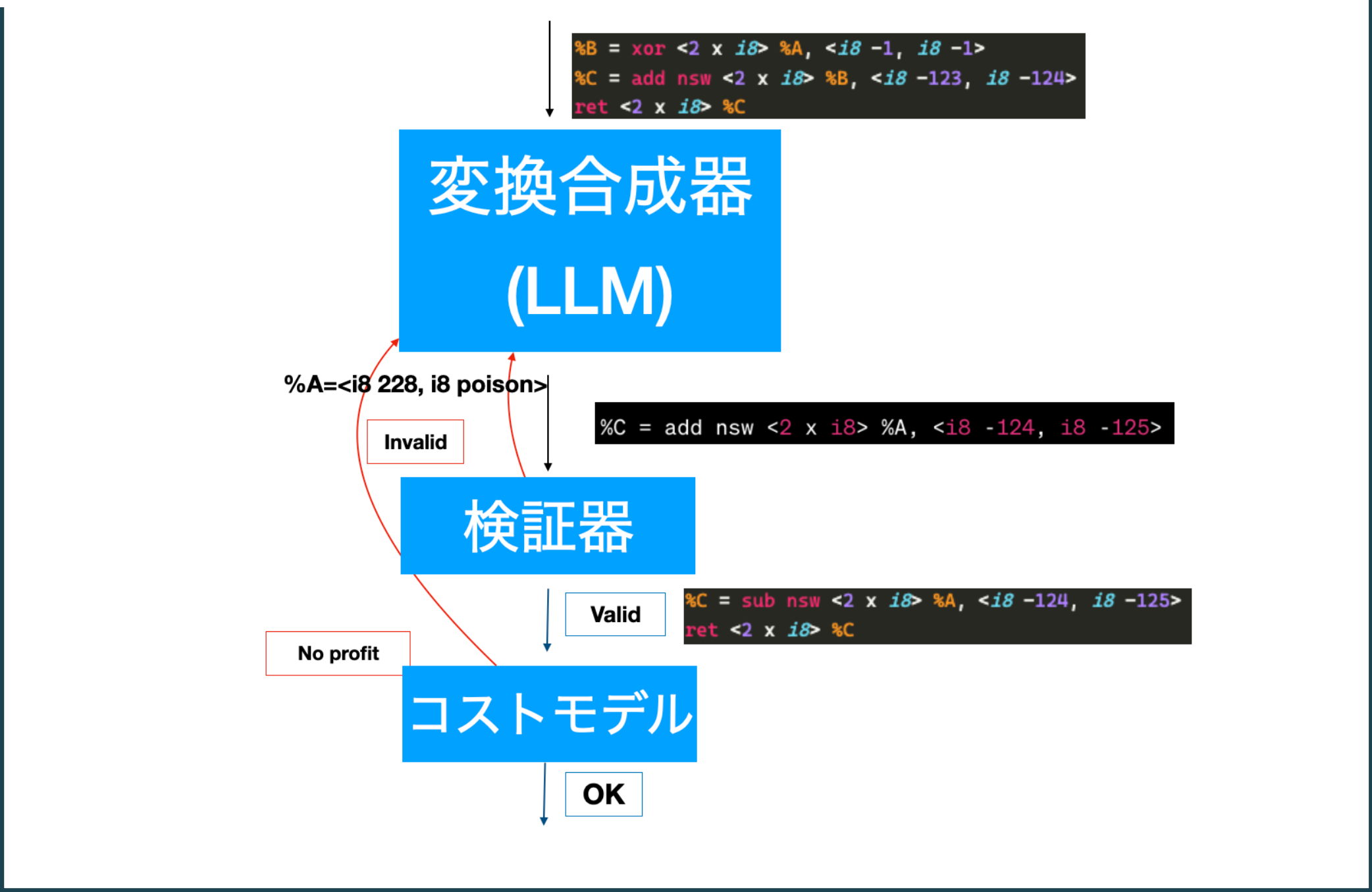
- プログラム片を元にして言語モデルによって最適化候補を作成
- 最適化検証器(alive2など)を用いて正当性を検証
 - Invalidな変換ならば反例を言語モデルにFeedback
- コストモデルの計算をし, 元のプログラム片よりも高速化が見込めるか検証(llvm-mca)
 - 高速化が見込めないなら言語モデルにFeedback
- プログラム片を置換

SIMD Superoptimizerの現状

- 合成した結果が必ずしも速いとはかぎらない
- コンパイル時間が爆発的

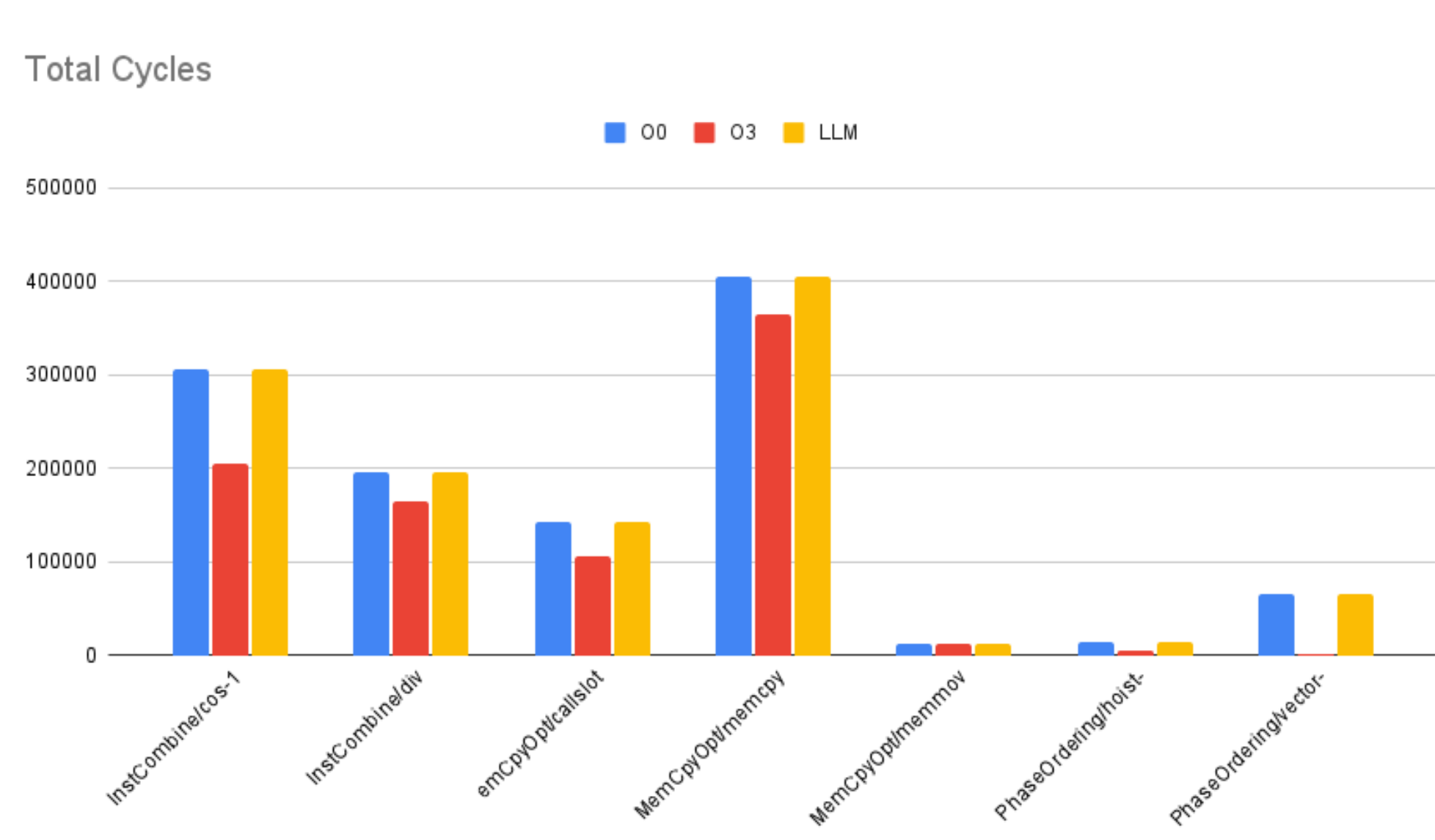


提案手法 概略



実験：llvm-projectの各種PASSのテストケース -O2/3との比較 (変換の正当性検証なし)

最適化パスのテストケースのCycle数比較



ARM SVE対応プロセッサのコストモデルで評価

まとめ

LLM baseのLLVM IR Super optimizer

- Fine tuning, Context learningをしないと意味のある変換は厳しい

今後の展望

- プログラムSlicingの利用
- 狙う変換を絞ったプロンプトの作成
- モデルのFine tuning, Context learningを利用したSIMD Intrinsicsのad hoc peephole 最適化

具体的なCycle数比較

testcase	-O0	-O3	LLM
InstCombine/cos-1	305901	204301	305901
InstCombine/div	195201	163901	195201
MemCpyOpt/callslot	142201	105702	142201
MemCpyOpt/memcpy	404102	365002	404102
MemCpyOpt/memmove	12901	13001	12901
PhaseOrdering/hoist-load-of-baseptr	14702	4501	14702
PhaseOrdering/vector-math	64901	701	64901

実際に使用したプロンプトとその返答

プロンプト	返答
<pre>Please optimize and/or vectorize following LLVM IR, if it's impossible, please return the provided code. PLEASE NEVER say anything except optimized LLVM IR PLEASE use opaque pointers and NEVER change function name. define i32 @add_zext_zext_i1(i1 %a) { %zext = zext i1 %a to i32 %add = add i32 %zext, %zext ret i32 %add }</pre>	<pre>define i32 @add_zext_zext_i1(i1 %a) { %zext = zext i1 %a to i32 %add = add i32 %zext, %zext ret i32 %add }</pre>