

XX-XX

卒業論文

LLVMコンパイラ基盤を用いた ベクトル化コード生成についての検討

2022年3月

永池 晃太郎

宇都宮大学工学部
情報工学科

内容梗概

日本語アブストラクト

Consideration of Vectorized Code Generation Using LLVM Compiler Infrastructure

Kotaro Nagaike

Abstract

ENGLISH abst

目次

内容梗概	i
Abstract	ii
目次	iii
1 はじめに	1
2 MIQS プロセッサ	3
2.1 RISC-V	3
2.2 MIQS 概要	4
2.3 ベクトル拡張付き RISC-V	5
3 LLVM	6
3.1 LLVM 概要	6
3.2 LLVM による自動ベクトル化機能	8
4 ベクトル拡張付き RISC-V コンパイラ	9
4.1 LLVM バックエンドにおける独自命令実装手法	9
4.2 ベクトル拡張付き RISC-V 命令の定義	10
5 評価と考察	12
5.1 生成されるアセンブリコードの評価	12
5.2 非対応命令についての考察	12
6 おわりに	13
謝辞	14
参考文献	15

第1章 はじめに

FPGA(Field Programmable Gate Array) はユーザによって回路の再構成が可能な LSI であり，目的の処理をハードウェアとして実装可能なデバイスである．近年 FPGA の大容量化，高性能化によって大規模な回路が実現可能になった．これにより FPGA は自動運転を始めとする組込み分野での利用増加が期待されている．

FPGA を用いたハードウェア開発は HDL(Hardware Description Language) による RTL(Register Transfer Level) 設計が広く用いられている．RTL は FPGA 上に構成する回路の信号の流れや制御構造を直接設計できる一方，動作検証やデバックが難しく，短期間で複雑な処理の開発は困難である．[1] そのため，FPGA による開発期間を短縮する方法として，専用ハードディスクとプロセッサを用いたソフトウェアによる処理を組み合わせる方法が考えられる．FPGA 上のハードウェアリソースを用いて実装するプロセッサのことをソフトコアプロセッサという．一般的に組込みシステムではそのコストやサイズ，消費電力などに制限があるためハードウェア資源に制約があることが多く，メモリバンド幅が限られることからメモリシステムの性能が高くないことが多い．メモリシステムの性能が低いと演算性能を高くしてもメモリアクセスに時間がかかり，結果としてシステム全体の性能はメモリシステムの性能によって左右される．[2] そこで，ソフトコアプロセッサによる処理を考える．すべての処理を専用ハードウェア回路で行うのではなく，高い性能の求められない汎用的な処理についてはソフトコアプロセッサにて行うことにより開発する専用ハードウェア回路を削減でき，ソフトコアプロセッサの性能が向上して専用ハードウェア回路を削減しても求められた性能に達することができればハードウェアの開発コストを抑えることができる．

組込み分野では AI 技術に注目が集まっている．独立行政法人情報処理推進機構の調査によると，将来強化/新たに獲得したい技術として組込み/IoT 関連企業の 46%が AI 技術を挙げている．[3]

AI 技術の応用としては画像認識などの画像処理が行われる．画像処理では画像を構成する画素に対して同じ処理を行うようなものが多い．このように複数のデータに対して同じ演算を行う処理については，単一命令で複数データの処理を行う SIMD(Single Instruction, Multiple Data) や，複数命令を並列に実行する MIMD(Multiple Instruction, Multiple Data) による並列処理で高速化が可能である．[4] MIMD は複数の制御を並列化することができるため，異なる処理を同時に実行するアプリケーションでは有効である．しかし，プロセッサに複数の制御ユニットをもたせる必要があるため SIMD

と比較するとハードウェアのコストが大きくなる．一方 SIMD は単一の制御ユニットで複数の演算ユニットを並列動作させるため制御ユニットのコストは低くなる．データ並列処理では複数のデータに対して同じ処理を実行するため SIMD によって並列処理が可能である．

SIMD による並列処理を行う場合，SIMD 命令は1命令で演算するデータ数が決まっている．そのため演算性能を上げるために同時演算数を変更すると機械語コードを作り直す必要がある．異なる同時演算数でも同一の機械語コードを利用可能とするためには，機械語コードが同時演算数に依存しないスケーラブルなベクトル拡張が必要である．スケーラブルなベクトル拡張により機械語コードを変更することなく，必要に応じて容易に同時演算数を増やし高性能化することが可能となる．

スケーラブルなベクトル拡張を実現したものとしてオープンな命令セットアーキテクチャである RISC-V[5] をベクトル拡張したベクトル拡張付き RISC-V[6] が提案されている．ベクトル拡張付き RISC-V は組込み機器に広く用いられている ARM のベクトル拡張である ARM SVE(Scalable Vector Extension)[7] の命令セットを参考に組み込み向けに RISC-V に拡張したものである．しかし，ベクトル拡張付き RISC-V に対応したコンパイラが存在していない．そこで解決策としてベクトル拡張付き RISC-V のベクトル命令のアセンブリコードを得るためのコンパイラの開発を検討した．

本論文では，第2章で現在の MIQS プロセッサについて述べる．第3章でベクトル拡張付き RISC-V 命令の生成のために利用したコンパイラ基盤である LLVM[8] について述べる．第4章で実際に命令生成のための実装について述べる．第5章では実際のソースコードからアセンブリコードの出力を行った結果について述べる．

第2章 MIQS プロセッサ

本章ではベクトル処理機能を持つソフトコアプロセッサである MIQS プロセッサについて述べる。

2.1 RISC-V

RISC-V はカルフォルニア大学バークレイ校が新たに開発した RISC の設計思想に基づく命令セットアーキテクチャ(ISA)である。RISC-V はオープンな ISA であり、ライセンスが不要である。

従来の ISA では、後方バイナリ互換性を維持するために過去に拡張した命令すべてを実装する必要がある。しかし、このような ISA では命令数が増加し複雑になる。そこで RISC-V ではモジュール式 ISA を採用している。[9] RISC-V のシステムの機能を独立したモジュールとして分け、アプリケーションに応じて拡張機能を組み込むかを選択できる。RISC-V には必ず組み込まなければならない基本命令セット (RV32I) の他に、主な拡張機能として乗算及び除算 (RV32M)、単精度浮動小数点 (RV32F)、倍精度浮動小数点 (RV32D) 等がある。

RISC-V の基本命令は 6 つの命令フォーマットで表すことができる。基本命令の命令フォーマットを図 2.1 に示す。

R 形式は 2 つのソースレジスタを扱う形式、I 形式は 12 ビットの即値を扱う命令、S 形式はストア命令、B 形式は条件分岐命令、U 形式は 20 ビットの即値を扱う命令、J 形式は無条件ジャンプ命令の形式である。命令形式が単純であるため命令のデコーディングが単純化される。

31	25 24	20 19	15 14	12 11	7 6	0	
funct7	rs2	rs1	funct3	rd	opcode		R-type
imm[11:0]		rs1	funct3	rd	opcode		I-type
imm[11:5]	rs2	rs1	funct3	imm[4:0]	opcode		S-type
imm[12:10:5]	rs2	rs1	funct3	imm[4:1:11]	opcode		B-type
imm[31:12]				rd	opcode		U-type
imm[20:10:1:11:19:12]				rd	opcode		J-type

図 2.1 基本命令の命令フォーマット

2.2 MIQS 概要

MIQS(MIPS Instruction processor with Quadword SIMD extension)[10] とは、既存の MIPS 命令互換プロセッサを元に SIMD 演算機能とオンチップメモリを備えたソフトコアプロセッサのことであったが、現在は RISC-V コアにベクトル演算機能をもたせたソフトコアプロセッサとなっており、特に SIMD 拡張を行っているわけではないが名称を引き継ぎ MIQS と呼称する。

図 2.2 に MIQS の全体構成を示す。MIQS の RISC-V コアプロセッサは既存の RISC-V プロセッサを使用している。このプロセッサは Verilog-HDL で記述された 5 段パイプラインプロセッサで回路の複雑化、規模増加を避けるために必要最低限の基本命令セットである RV32I に対応している。ベクトル処理ユニットは RISC-V 向けのベクトル命令セットに対応している。これによって機械語コードを同時演算数の変更によって作りなおす必要のないスケーラブルなベクトル拡張が実現できた。

命令のフェッチ、デコードは RISC-V コアプロセッサ上で行い、デコード結果がベクトル拡張命令であったときはベクトル処理ユニットで動作させる。

メモリアクセスはプロセッサからメモリコントローラを介して SDRAM にアクセスする。メモリアクセス要求がプロセッサから発行されたときは、メモリコントローラが SDRAM を制御してメモリアクセスを実現する。また、ベクトルメモリアクセスに関してもメモリコントローラにて行う。SDRAM コントローラのデータバス幅は 128 ビットであるから、プロセッサとメモリコントローラ間のデータバス幅は 128 ビットとなっている。

RISC-V コアプロセッサは命令フェッチ (IF)、命令デコード (ID)、実行 (EX)、メモリアクセス (MA)、ライトバック (WB) の 5 段のパイプラインで構成されている。ベクトル処理ユニットはベクトル長を 256 ビット、データの型は 32 ビット整数型を想定して設計されている。ベクトル処理ユニットと RISC-V コアの双方で動作する必要のある命令に対応するためベクトル処理ユニットは RISC-V コアと同じく 5 段パイプライン構成となっており、RISC-V コアと強調動作する構成となっている。なお、命令フェッチ部分と命令デコード部分は RISC-V コアと共通となっている。

MIQS によってスケーラブルなベクトル拡張を実現し、同時演算数の変更によって機械語コードの変更の必要がなくなったが、MIQS が対応している命令セットであるベクトル拡張付き RISC-V に対応しているコンパイラがないため、ベクトル化アセンブリコードを得ることができない。そのた

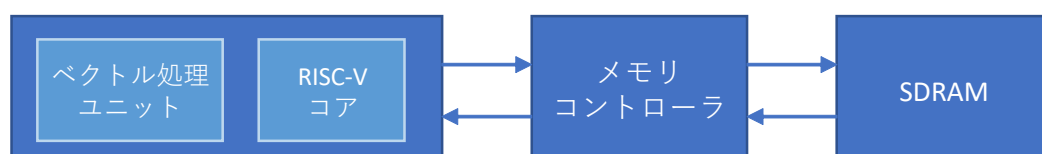


図 2.2 MIQS の構成

めコンパイラを作る必要がある。

2.3 ベクトル拡張付き RISC-V

ベクトル拡張付き RISC-V は機械語コードが同時演算数に依存しないスケーラブルなベクトル拡張を実現した ISA であり，既存の RISC-V を ARM SVE を参考にベクトル拡張したものである．ベクトル拡張付き RISC-V ではスケーラブルなベクトル拡張を行うためにプレディケートによるループ制御がある．プレディケートを用いたループ処理はループカウンタと処理する全データ数を比較してループカウンタが全データ数より小さい間対応するプレディケートレジスタを True にする．この命令によってベクトル処理で余りの要素がある場合，余りの要素の部分に対応するプレディケートの要素のみを True にする．これによって余りの要素が変化してもすべてのデータ数分ベクトル処理を行うことができ，同時演算数に依存しない機械語コードが実現できる．

ベクトル命令はベクトルロード，ストア命令，ベクトル演算命令，ベクトル制御命令の 3 つに分けられる．RISC-V にはカスタム命令用にオペコード領域が 4 つ用意されているがそのうちの 2 つを利用しており，1 つをベクトルロード，ストア命令，もう一方をベクトル演算命令とベクトル制御命令に使用している．ベクトルロード，ストア命令のアドレスはベースとなるスカラーレジスタの値にオフセットを加えることで計算する．ベクトル演算命令は，プレディケートあり演算命令，プレディケートなし演算命令，即値による演算命令に分けられる．浮動小数点命令は必要なハードウェア資源が多いためサポートしていない．ベクトル拡張付き RISC-V は基本命令に関して RV32I のみ組み込んでいる．

レジスタ構成はベクトルレジスタ $v0-v31$ ，ベクトルマスク制御に用いるためのプレディケートレジスタ $vp0-vp7$ ，プレディケートレジスタ同士の論理演算に用いるプレディケートレジスタ $vp8-vp15$ ，RISC-V の汎用レジスタ $x0-x31$ ，プログラムカウンタとなっている．ベクトルレジスタの長さは 128×2^n ($0 \leq n \leq 4$) ビットで表され，汎用レジスタの幅は 32 ビットとなっている．

命令のフォーマットは RISC-V の命令形式である R 形式に従ったものとなっている．また，デコーダの単純化のために同じフィールドにはできるだけ同じ機能をもたえている．

第3章 LLVM

本章ではベクトル命令拡張付き RISC-V のアセンブリコードを得るためのコンパイラの開発に用いたコンパイラ基盤である LLVM について述べる．

3.1 LLVM 概要

LLVM は 2000 年にイリノイ大学で開発が開始されたコンパイラ基盤である．コンパイラ基盤とはコンパイラに必要となるモジュールをまとめたもので，コンパイラを開発するためのフレームワークである．LLVM の構成を図 3.1 に示す．LLVM はソースコードを LLVM の中間表現である LLVM IR に変換するフロントエンド，LLVM IR に対して最適化等の操作や LLVM IR から機械語やアセンブリコードへの変換を行うバックエンドに分かれている．中間表現である LLVM IR は階層構造のドメイン固有言語である．LLVM ではこれらの機能がモジュール化されており，独自機能を実装する以外は既存のものを再利用することができる．例えば新たなアーキテクチャ向けのコンパイラを開発する際はバックエンドのみ実装を行い，フロントエンドについては再利用することができる．

LLVM には既に RISC-V を対象としたコード生成のためのバックエンドが実装されている．本研究では RISC-V を独自にベクトル拡張したベクトル拡張付き RISC-V の命令の生成を目的としているため，この RISC-V 向けのバックエンドに対して変更を加えることによって独自命令の生成を行う．

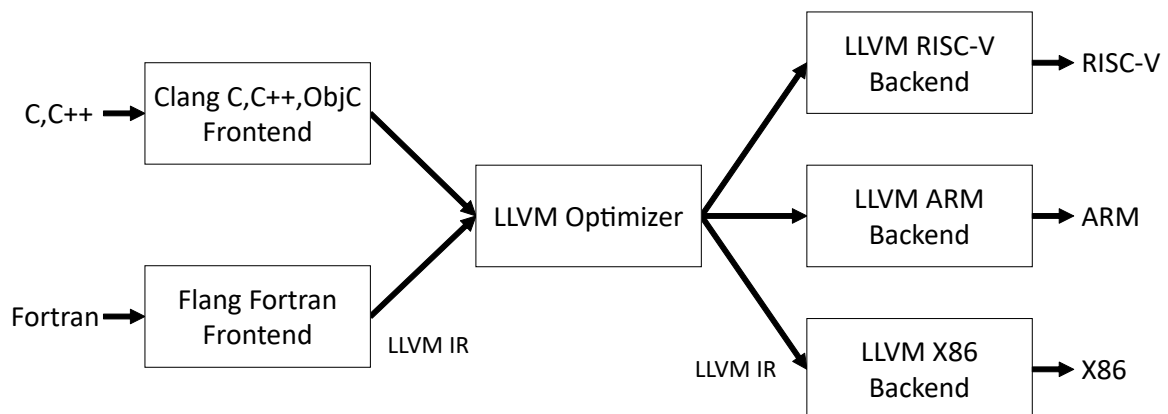


図 3.1 LLVM の構成

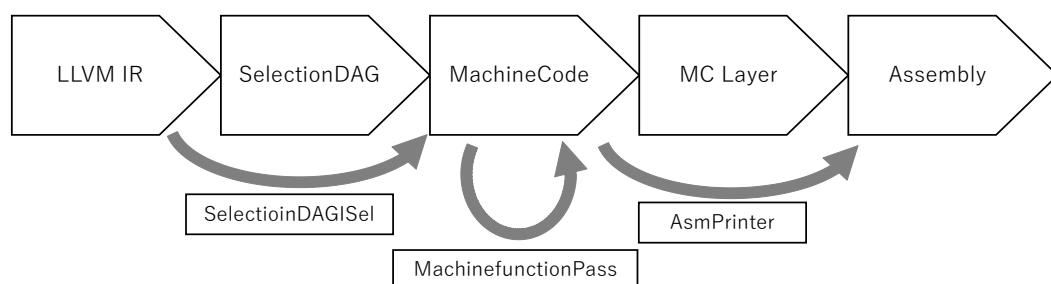


図 3.2 LLVM バックエンド

LLVM バックエンドにおけるコード生成の流れを図 3.2 に示す。LLVM バックエンドでは Pass によって処理が行われる。図 3.2 では LLVM バックエンドにおけるデータフォーマットの変化と実行される Pass を表している。

バックエンドでは LLVM IR から DAG(Directed Acyclic Graph) である SelectionDAG へフォーマットを変換する。SelectionDAG は LLVM IR をグラフ形式で表したもので、各命令やデータの依存関係を表現する。SelectionDAG への変換は SelectionDAGISel パスで行われ、SelectionDAG は最終的に SelectionDAGISel パスによって MachineCode 形式に変換される。SelectionDAGISel では Lower, Combine, Legalize, Select, Schedule のフェーズから構成される。

Lower は LLVM IR から SelectionDAG に変換させるフェーズである。このフェーズでは LLVM IR から SelectionDAG のノードへと一対一の対応を行う。この段階ではターゲットマシンでは利用できない命令やデータ形式を含んで不正 (illegal) な状態である。

Combine フェーズではパターンマッチングによる置き換えで最適化を行い、命令の単純化を行う。

Legalize はターゲットマシンではサポートされていない命令やデータ形式を他のものに置き換えるフェーズである。このフェーズによって不正な状態であった SelectionDAG ノードが正当 (legal) な状態となる。

Select フェーズでは SelectionDAG のノードをターゲットマシンの命令を含んだ MachineNode へと変換する。

Schedule フェーズでは構築されたグラフの依存関係を元に命令をスケジューリングするフェーズである。

SelectionDAGISel パスではこれらのすべてのフェーズが終わった後に MachineCode を出力する。

Machinefunction パスでは MachineCode という形式を扱う。この形式は LLVM IR と似た構成になっているが、より機械語に近い表現になっている。MachineCode が SelectionDAGISel によって生成された直後はまだ命令で扱うレジスタは無限個あると仮定した仮想レジスタや phi 関数を含んだ SSA(Static Single Assignment form) 形式で表現されている。MachinefunctionPass ではレジスタの割

当や phi 命令の削除を行い，MachineCode を非 SSA 形式へと変換する．

AsmPrinter パスは MachineCode を MC Layer 形式へと変換した後にアセンブリコードなどを出力するパスである．MC Layer 形式は MachineCode 形式のような階層構造がない形式であり，アセンブリコードへの変換だけでなく，オブジェクトファイルへの変換のための形式である．

3.2 LLVM による自動ベクトル化機能

LLVM には自動ベクトル化機能が備わっている．自動ベクトル化とは配列の演算処理などのループによる繰り返し処理をベクトル命令の形式に置き換える機能である．LLVM による自動ベクトル化機能はソースコードにおけるループをベクトル化された LLVM IR へと変換が行われる．LLVM による自動ベクトル化の例を図 3.3 に示す．LLVM IR においてベクトル命令はベクトル型を用いて表現される．図 3.3 にて `i128 x i32` となっている箇所がベクトル型の指定を行っており，これは 128 個の 32 ビット整数の演算を行うベクトル型の命令となっている．

また，LLVM IR では対象の全データに演算を行うための制御を行っている．例えば対象データ要素数が 400 とし，一度にベクトル演算を行う個数を 128 とする．この場合，128 個を対象としたベクトル演算を 3 回行う．するとベクトル演算では 384 個の要素を演算することができるが，16 個の要素が余っている．このあまりに関しては逐次処理によってスカラ演算を 16 回繰り返す処理を行う．これによってすべてのデータの演算を可能にしている．

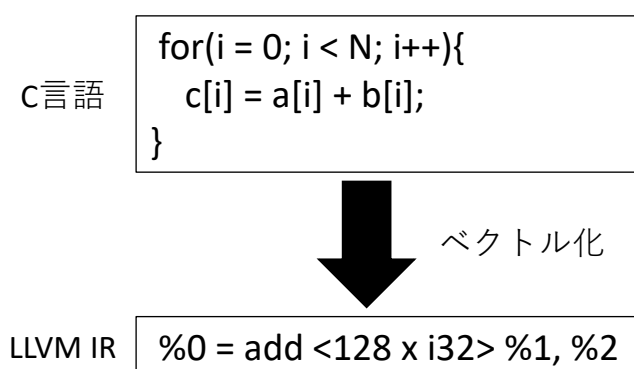


図 3.3 LLVM による自動ベクトル化例

第4章 ベクトル拡張付き RISC-V コンパイラ

本章ではベクトル拡張付き RISC-V のベクトル命令を LLVM によって生成するための命令実装手法と、実際に実装した命令の定義について述べる。

4.1 LLVM バックエンドにおける独自命令実装手法

LLVM において特定のターゲットマシン命令の生成は 3.1 で述べた SelectionDAGISel パスによる Select フェーズにて LLVM IR の命令からターゲットマシン命令に変換されることによって行われる。この変換は SelectionDAG のノードに対してパターンマッチングを行い、特定のパターンにターゲット命令を対応させる手法で行われる。

このパターンマッチングに用いるパターンは LLVM の固有ドメイン言語である TableGen によって行われる。TableGen はターゲットマシンの命令やレジスタ等の情報を記述するために用いられる。TableGen では命令生成のためにパターンの定義だけでなく、アセンブリコード等の出力のためにニーモニックの定義や命令フォーマットの定義も行っている。

SelectionDAG のパターンマッチングの例を図 4.1 に示す。図 4.1 では加算命令の例を示している。図 4.1 上にパターン定義クラス Pat がある。Pat クラスの第一引数と変換前の SelectionDAG のノードが一致した場合、Pat クラスの第二引数で指定したノードに変換する。

この様に SelectionDAG のパターンと命令ごとに定義されているパターンが一致したときに対応した命令にノードが変換される。LLVM では RISC-V の V 拡張命令のためのパターン定義が既に実装されており、そのパターンと一致した際に変換する命令を独自のベクトル拡張付き RISC-V の命令に定義しなおすことによってベクトル拡張付き RISC-V 命令の生成を実現する。

LLVM における命令の定義は命令フォーマットの定義と命令の定義に分かれる。命令フォーマットは TableGen によって命令の種類ごとにフォーマットのクラスを定義を行う。LLVM における命令フォーマットは基本クラス RVInst を継承する形で行われる。RVInst では 32 ビットのフィールド Inst や命令のニーモニックを格納する AsmString を定義している。この RVInst を継承して異なるフォーマットを定義していく。

命令の定義は命令フォーマットのクラスをインスタンス化する形で行われるが、そのために命令フォーマットを更に継承したクラスを定義する。このクラスは似たような種類の命令を定義するに

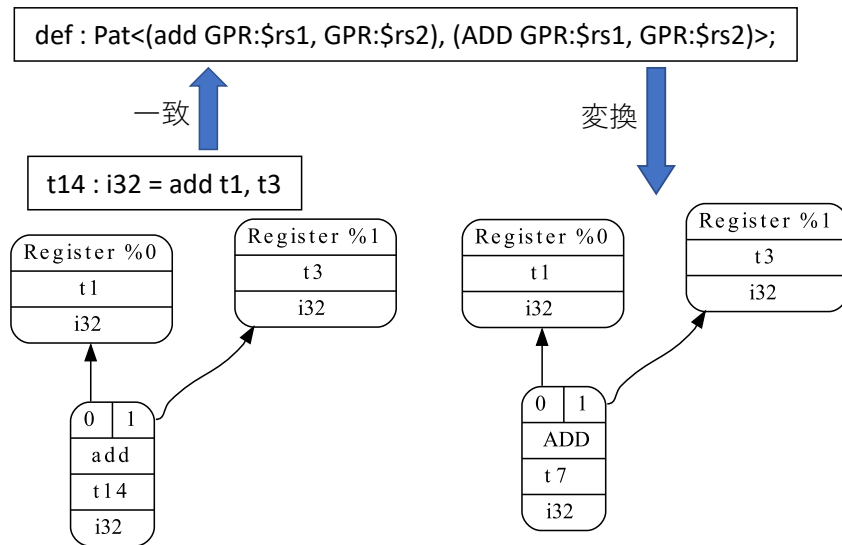


図 4.1 SelectionDAG の命令変換

当たって同じような定義を何度も繰り返さないように定義する。

命令フォーマット, 命令定義のクラスの継承関係を図 4.2 に示す。基本クラスである `RVInst` を継承して基本命令のフォーマットの定義を行っている。クラス `ALU_rr`, `ALU_ri` はそれぞれ算術演算命令のためのクラスである。

4.2 ベクトル拡張付き RISC-V 命令の定義

ここは 4.2 節です。

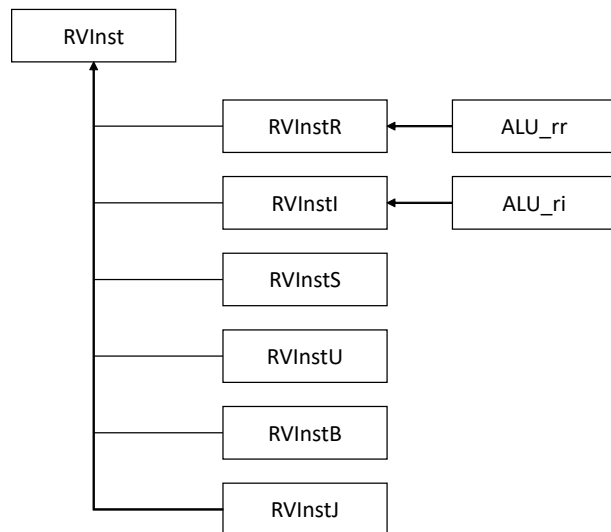


図 4.2 基本命令フォーマットクラス継承関係

第5章 評価と考察

5章．終盤，もう少し．

5.1 生成されるアセンブリコードの評価

ここは5.1節です．

5.2 非対応命令についての考察

ここは5.2節です．

第6章 おわりに

論文のまとめと課題．お疲れ様でした．

謝 辞

本研究の機会を与えていただき，また，日頃から貴重な御意見，御指導いただいた，馬場 敬信教授，大津 金光准教授，大川 猛助教，横田 隆史教授，に深く感謝致します．そして，本研究において多大な御力添えを頂いた，Boaz Jessie Jackin 氏をはじめとする研究室，オプティクス教育研究センターの方々に感謝致します．

参考文献

- [1] 三好健文: “FPGA 向けの高位合成言語と処理系の研究動向,” コンピュータソフトウェア. Vol.30, No.1, pp.76-84, 2013.
- [2] Samuel Williams, et al.: “Roofline: an insightful visual performance model for multicore architectures,” Communications of the ACM. Vol.52, No.4, pp.65-76, 2009.
- [3] 独立行政法人情報処理推進機構 社会基盤センター: “2020 年度組込み/IoT 産業の動向把握等に関する調査,” 2021.
- [4] デイビッド・A・パターソン, ジョン・L・ヘネシー著 成田 光彰 訳: “コンピュータの構成と設計 第 5 版,” 日経 BP 社. 2017.
- [5] Andrew Waterman, Krste Asanovi: “The RISC-V Instruction Set Manual Volume I: Unprivileged ISA, Document Version 2.2,” 2017.
- [6] Yoshiaki Kimura, et al: “Proposal of Scalable Vector Instruction Set for Embedded RISC-V Processor,” Proc. 2019 Seventh International Symposium on Computing and Networking Workshops (CANDARW), Vol.1, pp.435-439, 2019.
- [7] Nigel Stephens, et al: “The ARM Scalable Vector Extension,” IEEE Micro, Vol.37, No.2, pp.26-39, 2017.
- [8] Chris Lattner, Vikram Adve: “LLVM: A Compilation Framework for Lifelong Program Analysis Transformation,” Proc. 2004 International Symposium on Code Generation and Optimization (CGO '04), pp.75, 2004.
- [9] デイビッド・パターソン (著), アンドリュー・ウォーターマン (著), 成田 光彰 (訳): “RISC-V 原典オープンアーキテクチャのススメ,” 日経 BP 社, 2018.
- [10] 平石康祐, 橋本瑛大, 大津金光, 大川猛, 横田隆史: “SIMD 拡張ソフトコアプロセッサのための効率的なメモリシステムの検討,” 情報処理学会第 80 回全国大会講演論文集, pp.115-116, 2018.