

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

永池 晃太郎

1 はじめに

画像認識等のデータ並列性の高い処理の高速化には SIMD 命令による処理が効果的である。しかし、SIMD 命令による処理は同時演算数を変更する場合、同時に機械語コードを作り直す必要がある。そこで機械語コードの変更なしに同時並列演算数を変更することができる、データ並列処理のためのスケーラブルなベクトル処理機能を備えたベクトル拡張付き RISC-V を開発された^[1]が、これに対応したコンパイラがない。

本研究では、既存の RISC-V コンパイラをベースとして、ベクトル拡張付き RISC-V に対応した自動ベクトル化コンパイラを開発する手段を検討する^[3]。

2 LLVM コンパイラ基盤

コンパイラの開発にはコンパイラ基盤である LLVM^[2]を用いる。コンパイラ基盤はコンパイラの各機能がモジュール化されており、機能の再利用が行いやすくなっている。そのため、LLVM で提供されている各種モジュールを利用することで、独自機能部分の実装に集中することができる。また、LLVM には自動ベクトル化機能が備わっており、入力ソースコードの繰り返し処理をベクトル化された LLVM IR に変換する。ベクトル化された LLVM IR では、処理対象の全データへ演算を行うために複数データに対してベクトル演算を繰り返した後に残りのデータ要素がある場合は逐次処理によって演算を行う。その LLVM IR からパターンマッチングにより各ターゲットへのベクトル命令への変換などが行われる。ベクトル拡張付き RISC-V は RISC-V を拡張しているため、LLVM の RISC-V コンパイラとしての機能をベースとしてコンパイラを開発を行う。

なお、使用する LLVM のバージョンは 13.0.0 である。

LLVM はソースコードから中間表現である LLVM IR に変換を行うフロントエンド、LLVM IR からアセンブリコード生成を行うバックエンドからなる。

LLVM のバックエンドでは LLVM におけるターゲットマシン情報記述のためのドメイン固有言語である TableGen を用いて記述される、レジスタ情報、命令フォーマットや命令のニーモニック等の定義に従って LLVM IR からアセンブリコードへの変換を行っている。本研究では独自命令生成のためにバックエンドに対して変更を加える。

3 LLVM における独自命令の生成機能の実装

ベクトル拡張付き RISC-V 命令の定義のために、各種命令用のフォーマットクラスの定義を行う。図 1 はベクトル要素同士を演算する命令であるベクトル加算等の命令用のフォーマットの定義である。命令フィールド Inst のどの範囲にどの値が入るかを指定している。

```
1 :  
2 let Inst{31-27} = 0b00001;  
3 let Inst{26-25} = 0b10;  
4 let Inst{24-20} = vs2;  
5 let Inst{19-15} = vs1;  
6 let Inst{14-12} = funct3;  
7 let Inst{11-7} = vd;  
8 let Inst{6-0} = 0b00010111;  
9 :
```

図 1: 命令フォーマットの定義

```
class VALUVVMIQS<bits<3> funct3, string opcodestr>  
: RVInstVVMISQ<funct3, (outs VR:$vd), (ins VR:$vs2, VR:$vs1),  
opcodestr, "$vd, $vs2, $vs1"> {  
}  
def VADD_VV : VALUVVMIQS<0b000, "vadd.w">,  
Sched<[WriteVIALUV, ReadVIALUV, ReadVIALUV]>;
```

図 2: 命令の定義

```
void add(int *a, int *b, int *c){  
for (int i=0;i<N;i++){  
c[i] = a[i] + b[i];  
}  
}
```

⋮		
vle32.v	v25, (a7)	#ベクトルロード
vle32.v	v26, (a5)	#ベクトルロード
vadd.w	v25, v26, v25	#ベクトル加算
vse32.v	v25, (a3)	#ベクトルストア
⋮		

図 3: アセンブリコードの生成

命令の定義は命令定義用クラスをインスタンス化する際に命令文字列や命令選択のための命令フィールドの値の指定で行う。図 2 にベクトル加算命令の定義を示す。図 2 の VALUVVMIQS クラスではニーモニックの構成を指定しており、インスタンス化を行う際に命令文字列である vadd.w と命令選択のための値の指定を行っている。

配列加算のプログラムから実際にアセンブリコードを生成した結果の一部を図 1 に示す。vadd.w がベクトル拡張付き RISC-V 命令のベクトル加算命令である。しかし、ロード・ストア命令については RISC-V の V 拡張の命令のみである。

現在ベクトル拡張付き RISC-V のベクトル命令の内、ベクトル演算命令のプレディケートなし、即値による演算命令の実装が完了している。しかし、ロード・ストア命令等が未実装である。それらの命令ではプレディケートレジスタを用いており、新たにプレディケートレジスタの定義を必要とする。また、現在の LLVM IR はベクトル演算の繰り返しと逐次処理によるベクトル処理を行っている。そのため、ベクトル拡張付き RISC-V のプレディケート付き命令を効果的に利用できない。

4 おわりに

本研究では、自動でベクトル化された独自のベクトル拡張付き RISC-V 命令アセンブリコードを得るための LLVM バックエンドの実装を行った。また、実際に C 言語で記述した配列加算プログラムからアセンブリコードの生成を確認した。

今後の課題として、ベクトルロード・ストア命令等の未実装の命令生成の実現が挙げられる。また、新たにレジスタの定義に加え、LLVM IR の生成を行うフロントエンドの変更が必要である。

参考文献

- [1] Yoshiki 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.
- [2] 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.75-86, 2004.
- [3] 永池 晃太郎, ほか: “独自ベクトル処理機能を備えたプロセッサ向け自動ベクトル化コンパイラの開発,” 情報処理学会 第 84 回全国大会 講演論文集, 2022.(発表予定)