

PIKG仕様書

Ver 0.0 — 2020/7/31

野村昂太郎

August 12, 2020

Contents

1	TODO	3
2	履歴	3
2.1	2019/12/23	3
2.2	2019/12/30	3
2.3	2020/5/7	3
2.4	2020/7/31	3
2.5	2020/8/7	3
3	はじめに	3
4	概要	3
5	DSL仕様	4
5.1	粒子間相互作用カーネルの一般化	4
5.2	DSL概要	4
5.3	文法等	5
5.3.1	コメントアウト	5
5.3.2	予約語とその説明	5
5.3.3	利用可能な型	5
5.3.4	演算子	6
5.3.5	条件分岐	6
5.3.6	予約関数	7
5.4	変数宣言部	7
5.5	関数記述部	8
5.6	相互作用関数記述部	8
5.6.1	プラグマ記述	8
6	カーネル生成方法及び利用法	8
6.1	生成されるカーネルの形式	9

1 TODO

- FXXmat関連の演算や型推論が実装されていない
- 現状勝手にアップキャストされてしまうが、低精度を維持したい場合、statementの先頭に型宣言ができる必要がある
- 16 bit型の演算をサポートする

2 履歴

2.1 2019/12/23

ファイルを作った.

2.2 2019/12/30

カーネルジェネレータ周りのTODOを追加.

2.3 2020/5/7

カーネルジェネレータの名前をPIKGに変更.

2.4 2020/7/31

0.0版リリース向けに改定

2.5 2020/8/7

変数宣言部に記述を追加.

2.6 2020/8/12

条件分岐の文法に間違いがあったので修正.

3 はじめに

本書は粒子間相互作用カーネル関数のソースコードジェネレータPIKGの仕様書である.

粒子間相互作用とは、分子動力学やN体、SPHなどの粒子系シミュレーションにおいて、粒子の時間発展を差分して解く際に計算する粒子間に働く相互作用のことである. ここでは、2体間相互作用のみを考える.

粒子系シミュレーションでは、一般的に計算量のほとんどをこの粒子間相互作用の計算が占めている. したがって、多様化しているさまざまな計算機上で十分な実行性能を引き出すためには、それぞれのアーキテクチャ上で相互作用を計算するカーネルの最適化が必要不可欠である. しかしながら、これらの最適化カーネルは特定のアーキテクチャに特化した形で書かれるものが多く、計算機を変える場合には新たにカーネルを書き直すことがほとんどである. さまざまな計算機を利用する研究者にとって、それぞれのアーキテクチャ向けにカーネルを最適化することは、多くの時間を要するだけでなく、複雑なアーキテクチャの特性を理解しないといけないためそもそも最適化が困難であるという問題がある.

相互作用計算カーネルジェネレータPIKGは、ひとつのDSL(ドメイン特化型言語)記述から、複数のアーキテクチャ向けに最適化された相互作用計算カーネルを生成することを目的としている.

4 概要

カーネル関数の最適化ソースコードを様々なアーキテクチャ向けに生成するためのジェネレータを作る。そのための言語仕様及びコンパイラの仕様、簡単な使い方を示す。

5 DSL仕様

PIKGは、PIKG用DSL(Domain Specific Language)で書かれたソースコードを元に、指定したアーキテクチャ向けに最適化されたC++のカーネルソースコードを生成する。

以下では本カーネルジェネレータで用いるDSLの仕様について定義する。

5.1 粒子間相互作用カーネルの一般化

まず、DSL記述について述べる前に、粒子間相互作用カーネルが一般にどのような構造をとるのかを説明する。C/C++の関数では2粒子間相互作用カーネルは以下のように書ける。

```
void kernel_body(const EPI* epi,const int ni,
                const EPJ* epj,const int nj,
                FORCE* force){
    (1) preprocess
    for (int i=0; i < ni; i++){
        (2) load EPI variables
        (3) load FORCE variables
        for(int j=0; j < nj; j++){
            (4) load EPJ variables
            (5) calc interaction for FORCE variable
        }
        (6) store FORCE variables
    }
}
```

入力として、相互作用をうける粒子であるEPIと相互作用を及ぼす粒子であるEPJのArray of Structure(AoS)形式の配列epiとepj(これらをあわせて相互作用リストと呼ぶ)、それぞれの配列長niとnj、計算された相互作用を保持する構造体であるFORCEのAoS形式の配列が与えられる(FORCEの配列の長さはni)。

(1)ではEPIやEPJの変数を一時配列に入れ直し、AoSからSoA形式にする等の前処理が行われる。(2)ではEPI変数及び(1)で作った一時配列のうち、カーネル計算に必要な変数のみがロードされる。(3)では相互作用のアクキュレートに必要なFORCE変数をロードする。FORCE変数は事前に初期化されていることが前提となっている。これは、複数のカーネルを使ってFORCEを計算する場合を考慮しているためである。(4)ではEPJ変数をロードする。(2)と同様である。(5)ではロードした変数を使って相互作用の計算を行い、FORCE変数にアクキュレートしていく。最後に(6)でアクキュレートしたFORCE変数をもとの配列にストアする。

PIKGでは、DSL記述から(5)相互作用計算と(5)で使われる変数がEPI、EPJとFORCEクラスのどの変数かの情報を読み取り、相互作用計算カーネルを生成する。

5.2 DSL概要

PIKGのDSLは以下の3つのパートから構成される。

- 変数宣言部
- 関数宣言部
- 相互作用記述部

それぞれのパートは上記の順番で書かれる必要があり、関数宣言部は省略が可能である。変数宣言部はひとつ以上の変数宣言からなり、関数宣言もひとつ以上の関数宣言からなる。相互作用記述部では、変数宣言部と関数宣言部で宣言された変数と関数、予約関数を用いて粒子間の相互作用を記述する。

以下では、DSLの文法とそれぞれのパートについて詳しく見ていく。

5.3 文法等

5.3.1 コメントアウト

二重のスラッシュ(//)以降のその行の文字列はコメントとして無視される。C++における/* */のような行をまたげるコメントアウトは存在しない。

5.3.2 予約語とその説明

- EPI (変数宣言時に相互作用を受ける粒子の変数であることを示す)
- EPJ (変数宣言時に相互作用を及ぼす粒子の変数であることを示す)
- SPJ (変数宣言時に相互作用を及ぼす粒子でTree法で見込角より外のまとめられた粒子であることを示す, MultiWalkのときのみ使用)
- FORCE (変数宣言時に求める相互作用を保持する変数であることを示す)
- local (変数宣言時にlocalメンバ変数であることを示す)
- function (関数宣言であることを示す)
- (S|U)(16|32|64) ((符号あり|なし)整数スカラー型)
- F(16|32|64)([vec|mat][2|3]) (浮動小数点スカラー, ベクトル, マトリックス型. 末尾の数字は次元. 何もつけない場合は3次元として扱う)
- (sqrt|rsqrt|inv|min|max|table|to_float|to_int|to_uint|madd|msub|nmadd|nmsub) (特殊関数. それぞれ平方根, 逆数平方根, 逆数, 最小値, 最大値, テーブル関数, 積和算)

5.3.3 利用可能な型

本DSLでは、符号あり/なし整数型および浮動小数点スカラー型、ベクトル型、マトリックス型をサポートする。具体的には、以下の型が使用可能である。(マトリックス型は対称行列のみだが現状サポート外)。

- (S|U)(16|32|64)
- F(16|32|64)
- F(16|32|64)vec(2|3|4)?
- F(16|32|64)mat(2|3|4)?

それぞれ、16 bit, 32 bit, 64 bitの(符号あり | なし)整数型, 浮動小数点スカラー型, 浮動小数点ベクトル型, 浮動小数点マトリックス型である。浮動小数点ベクトル型とマトリックス型は末尾に2もしくは3, 4をつけることにより、次元を表すことができる。省略した場合は3次元である。

仕様に定めているが現状未サポートの機能は以下。

- 現状16 bit型の変数
- すべてのマトリックス型

5.3.4 演算子

使用可能な演算子は以下である.

- + (加算)
- - (減算)
- * (乗算, ベクトル型は内積)
- / (除算)
- && (論理積)
- || (論理和)
- == (等価)
- != (不等価)
- < (比較)
- > (比較)
- <= (比較)
- >= (比較)
- += (加算代入)
- -= (減算代入)
- *= (乗算代入)
- /= (除算代入)

詳細は相互作用関数記述部5.6節で説明する.

5.3.5 条件分岐

条件分岐は以下の文法でおこなえる.

```
if 条件1
  実行1
elsif 条件2
  実行2
else
  実行3
endif
```

elsif以下は省略可能である. 将来的にendifはendでも使えるようになる予定(Rubyの文法に合わせる). 後置などではない.

例

```
dr = ri - rj
r2 = dr*dr
if r2 > 0.0 && r2 < 16.0
  r2i = inv(r2)
  r6i = r2i * r2i * r2i
  f += 24 * r6i * r2i * (2*r6i - 1) * dr
  u += 4 * r6i * (r6i - 1)
endif
```

5.3.6 予約関数

予約されている関数は以下.

- sqrt (平方乗根)
- rsqrt (逆数平方根)
- max (最大値)
- min (最小値)
- madd ($A * B + C$)
- msub ($A * B - C$)
- nmadd ($-A * B - C$)
- nmsub ($-A * B + C$)
- table (テーブル命令)
- to_(float|int|uint) (キャスト)

積和算系の命令に関しては、DSL中で使わなくても、自動的に縮約される.

5.4 変数宣言部

変数宣言では、型の前にEPI, EPJもしくはFORCEをつけることでそれぞれのメンバ変数であることを示す. また、変数宣言では変数名のあとにコロン(:)と実際のEPIもしくはEPJ, FORCE型のメンバ変数名を書かなくてはならない. これにより、変数宣言で宣言された変数名と実際の型のメンバ変数の対応関係を示す.

```
EPI F32vec ri:r
EPJ F32vec rj:r
EPJ F32    mj:m
FORCE F32 fi:f
FORCE F32 ui:u
```

これは仕様ではないが、現状"x","y","z","w"は変数名として使用することができない.

EPI, EPJもしくはFORCEがつかない場合、カーネル関数オブジェクト初期化時に渡す引数として扱われる. DSL上ではグローバル変数のような扱いになる. この場合、コロンによって対応関係を示す必要はない.

F32 eps

また、local修飾子をつけることでEPI, EPJのローカル変数を宣言できる. ローカル変数にEPIやEPJの変数を使った値を代入するDSLを記述すると、相互作用計算の2重ループの外側で初期化される. local変数は1次元配列として確保されるため、EPIやEPJの要素数が大きいときなどにAoSからSoAに変換するときに用いる.

EPI, EPJ, FORCEの変数宣言はC++のクラスの宣言と型やメンバ変数の宣言の順番、個数と完全に一致してはならない. したがって、実際には相互作用の計算に使用されないメンバ変数がクラスに含まれている場合でも、変数宣言部では宣言される必要がある.

5.5 関数記述部

本DSLでは、関数を記述し、相互作用関数記述部で利用することができる。関数記述部は、以下の記述になる。

```
function 関数名(引数0, $\cdots$)
  /* 計算部分 */
  return 返り値
end
```

関数記述内の計算部分では、予約語および変数宣言部のない任意のローカル変数を使用できる。型推論されるので、特に宣言なく使用できる。

5.6 相互作用関数記述部

相互作用関数記述部では、以下のような記述が可能である。

```
dr = ri - rj
r2 = dr * dr
rinv = rsqrt(r2)
f += mj * ri * ri * ri * dr
u += mj * ri
```

相互作用関数記述部では、これまでに型を指定して宣言された変数をもとに型推論が行われるため、ローカルな変数の宣言などを行う必要がなく、これまでに宣言されていない好きな変数名の変数に値を代入できる。

カーネルジェネレータは、EPIとEPJの2重ループを作り、その中でこれらの演算子を用いて書かれた記述をアキュムレートするカーネル関数を生成する。

5.6.1 プラグマ記述

主にARM SVE向けの機能であるが、プラグマをはさみたい場所に`#pragma`から始まる行を入れると、生成されたカーネルの対応する箇所と同じプラグマ文が入る。以下のプラグマは、実際に生成されるカーネルにも影響を与える。

- `#pragma unroll X` (Xはアンロール段数)
- `#pragma statement loop_fission_point`

6 カーネル生成方法及び利用法

カーネルの生成方法は簡単である。以下の2ステップで行われる。

- 相互作用関数をDSLで記述
- オプションを指定してカーネルジェネレータにDSLファイルを入力として渡す

指定可能なオプションは以下の通りである。

- `--input | -i` ファイル名 (入力ファイル名の指定)
- `--output | -o` ファイル名 (出力ファイル名の指定, デフォルトは`kernel.hpp`)
- `--conversion-type` (AVX2|AVX-512|SVE) (アーキテクチャの指定)
- `--kernel-name` カーネル関数名 (カーネル関数名の指定. デフォルトは`Kernel`)

- `--epi-name` EPI名 (EPIの構造体名の指定, デフォルトはEPI)
- `--epj-name` EPJ名 (EPJの構造体名の指定, デフォルトはEPJ)
- `--force-name` FORCE名 (FORCEの構造体名の指定, デフォルトはFORCE)
- `--strip-mining` ストリップマイニンググループ数 (アーキテクチャがSVEに指定されている場合にストリップマイニングを行う)
- `--unroll` アンロール段数 (`--strip-mining`指定時に最内ループをアンロールする段数を指定, デフォルトは1. DSL内に`unroll`のプラグマが現れるとそれ以降はプラグマで指定されたアンロール段数が使われる)
- `--multiwalk` (FDPSにおけるMultiwalk向けにカーネルを生成する場合に指定, 現状利用不可)
- `--spj-name` SPJ名 (SPJの構造体名の指定. `--multiwalk`を指定したときのみ利用される. デフォルトはSPJ)

`--strip-mining`及び`--unroll`は将来的に廃止される予定であるが, 現状ARM SVE向けの最適化には指定が必要なので公開する. 現状, 最適なARM SVEカーネルを生成するためには, DSL内の適切な箇所に`unroll`及び, `loop_fission_point`のプラグマを挿入する必要がある.

6.1 生成されるカーネルの形式

カーネルは以下のような構造体として生成される(いずれもデフォルトのEPI,EPJ,FORCE, カーネル名を使う場合).

```
struct Kernel{
    void operator()(const EPI* epi, const int nepi, const EPJ* epj,const int nepj,FORCE* force){
        (カーネル本体)
    }
};
```

生成したカーネルは生成されたファイル(ここでは`kernel.hpp`)をインクルードし,

```
#include "kernel.hpp"
const int N = 1024; // 粒子数
int main{
    const int nepi = N;
    const int nepj = N;
    EPI *epi = new EPI[nepi];
    EPJ *epj = new EPJ[nepj];
    FORCE *force = new FORCE[nepi];
    (epiやepjに粒子の情報を代入)
    Kernel(epi,nepi,epj,nepj,force); // カーネル呼び出し
    (後略)
}
```

のようにして利用できる.