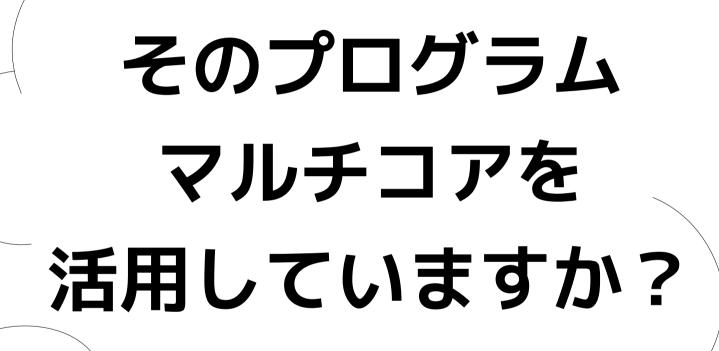
新しい並列for構文のご提案

2014/3/1 Boost.勉強会 #14 東京



何の変哲も無いプログラム

```
#include <vector>
using namespace std;
const int N = 1000000;
vector<MyData> data = /* N要素 */;
for (int i = 0; i < N; ++i) {
   process(data[i]);
```

何の変哲も無いプログラム

forループの処理が重い…

```
vector<myData> uata = / n女术 /,
```

```
for (int i = 0; i < N; ++i) {
    process(data[i]);
}</pre>
```

何の変哲も無いプログラム

forループの処理が重い… そうだ、**並列処理**しよう

```
VECTOI < MYData > Uata = / N女术 /,
```

```
for (int i = 0; i < N; ++i) {
    process(data[i]);
}</pre>
```

並列処理

↓ マルチスレッド ↓

Pthreads, WinAPI, etc.

並列処理

え? もう2014年ですよ?

Pthreads, WinAPI, etc.

並列処理

↓ マルチスレッド ↓

C++11 std::thread

or Boost.Thread

並列化 powered by C++11

```
int m = max(thread::hardware_concurrency(), 1u);
vector<thread> worker:
for (int i = 0; i < m; ++i) {
   worker.emplace_back([&](int id) {
      int r0 = N/m * id + min(N%m, id);
      int r1 = N/m * (id+1) + min(N%m, id+1);
      for (int j = r0; j < r1; ++j) {
         process(data[j]);
   }, i);
for (auto& t : worker) t.join();
```

並列化 powered by C++11

```
int m = max(thread::hardware_concurrency(), 1u);
vector<thread> worker:
                                       プロセッサ数取得
for (int i = 0; i < m; ++i) {
                                      スレッド牛成
   worker.emplace_back([&](int id) {
      int r0 = N/m * id + min(N%m, id);
      int r1 = N/m * (id+1) + min(N%m, id+1);
     for (int j = r0; j < r1; ++j) {
                                          区間均等分割
        process(data[j]);
   }, i);
                        forループ処理
for (auto& t : worker) t.join();
                                      スレッド終了待機
```

並列化 powered by C++11

並列化のためのコードでなく ロジック記述に注力したい

```
int r1 = N/m * (id+1) + min(N%m, id+1);
for (int j = r0; j < r1; ++j) {
    process(data[j]);
    }
}, i);

for (auto& t : worker) t.join();</pre>
```

ループの並列化といえば

ループの並列化といえば



並列化 powered by OpenMP

```
// 逐次処理
for (int i = 0; i < N; ++i) {
  process(data[i]);
                           // OpenMP並列化
                           #pragma omp parallel for
                           for (int i = 0; i < N; ++i) {
                              process(data[i]);
```

並列化 powered by OpenMP

どのように並列処理するか?

→ なにを並列処理するか?

手続的(HOW)な記述

→ **宣言的(WHAT)**な記述

おしまい





C++時代の

新しい並列for構文のご提案

2014/3/1 Boost.勉強会 #14 東京

あらためまして

誰?

twitter @yohhoy / hatena id:yohhoy



何を?

C++と親和性の高いループ並列化技術の紹介

どうして?

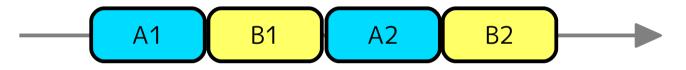
並列処理とか流行ってるから

C++11以降、並行・並列処理の検討が活発化

並行と並列

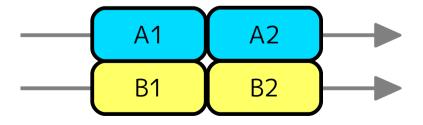
並行(Concurrent)

複数のタスクが存在し、それぞれが実行される。 "本当に同時刻に処理されるかどうか"は気にしない。



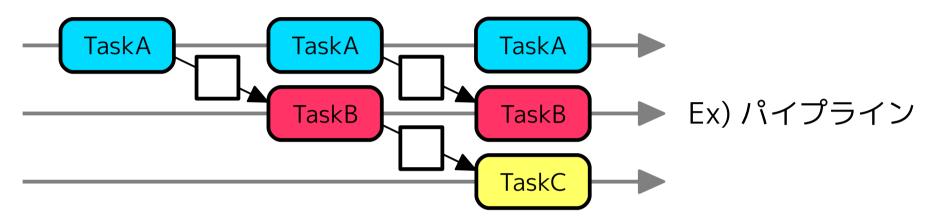
並列(Parallel)

高速化を目的として、複数H/Wを用いた同時計算を行う。 "速く"ならなければ無価値。

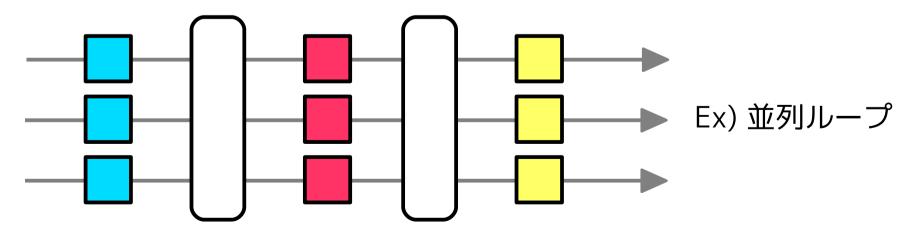


2種類の並列性

タスク並列性(Task Parallelism)



データ並列性(Data Parallelism)



並列化のレベル

大規模

マシンレベル並列(分散処理)

プロセスレベル並列

スレッドレベル並列

ベクトルレベル並列(SIMD)

命令レベル並列

マシンの壁

共有/分散メモリの壁

データモデルの壁

自動化の壁

高密度

(※今回の内容に合わせた独自の分類)

コンパイラによる 完全な自動並列化は 「人類の夢」

Fully automatic parallelization is still a

"pipe dream"

【可算名詞】 〔アヘン吸引者が見るような〕 夢想、幻想

Parallel and Concurrent Programming in Haskell, O'REILLY, 2013

Real World Parallel Programming...

ライブラリ/言語拡張を利用した 並列処理プログラミングが必要

マシンレベル並列(分散処理)

プロセスレベル並列

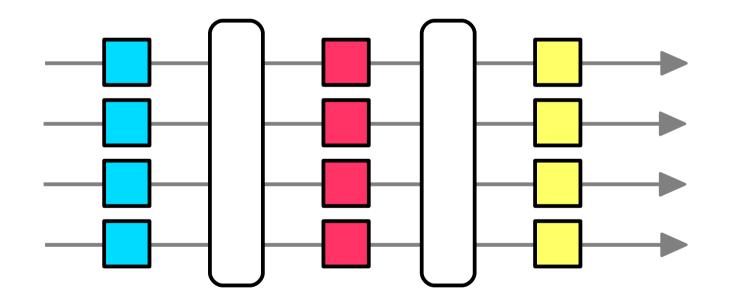
スレッドレベル並列

ベクトルレベル並列(SIMD)

命令レベル並列

スレッドレベル並列

(マルチスレッド)



スレッドレベル並列を支える技術

```
for (int i = 0; i < N; ++i) {
    process(data[i]);
}</pre>
```

ライブラリ

TBB, PPL, libstdc++ ParallelMode

言語拡張

CilkPlus, OpenMP

OpenMP

#pragma omp parallel for

```
for (int i = 0; i < N; ++i) {
    process(data[i]); // C++例外をthrow...
}
```

対応コンパイラ: GCC, Clang, MSVC, ICC...

OpenMP

```
try {
  #pragma omp parallel for
  for (int i = 0; i < N; ++i) {
    process(data[i]); // C++例外をthrow...
  } // parallel regionここまで
} catch (...) {
                       OpenMPスレッドをまたぐ
  // 例外catch ??
                          C++例外送出はNG
```

対応コンパイラ: GCC, Clang, MSVC, ICC...

OpenMPとC++例外

```
bool failed = false;
#pragma omp parallel for
for (int i = 0; i < N; ++i) {
   try {
       process(data[i]); // C++例外をthrow
   } catch (...) {
       // 同一スレッド内でcatch
       failed = true;
} // parallel regionここまで
if (failed) {
   // ...
```

C++例外の扱いがめんどう (C++的でなく**ダサい**)

> 例外throwされても ループ中断できない

bug free?

OpenMPとC++例外

```
bool failed = false;
                                                データ競合
#pragma omp parallel for reduction(||:failed)
                                               問題を修正
for (int i = 0; i < N; ++i) {
   try {
      process(data[i]); // C++例外をthrow
   } catch (...) {
                                   コード複雑化による
      // 同一スレッド内でcatch
      failed = true;
                                    厄介なバグの温床
} // parallel regionここまで
if (failed) {
```

// ...

libstdc++ Parallel Mode拡張

```
// g++ soruce.cpp -D_GLIBCXX_PARALLEL -fopenmp
#include <algorithm>
std::for_each(begin(data), end(data),

[](Data& item) {
    process(item);
    // 例外送出は禁止!
});
```

対応コンパイラ: GCC

対応OS: GCC+OpenMPが動く環境

libstdc++ Parallel Mode拡張

```
#include <parallel/algorithm>
__gnu_parallel::for_each(begin(data), end(data),

[](Data& item) {
    process(item);
    // 例外送出は禁止!

});
```

対応コンパイラ: GCC

対応OS: GCC+OpenMPが動く環境

Threading Building Blocks

Intel TBB

```
#include <tbb.h>
  try {
     tbb::parallel_for_each(begin(data), end(data),
     [](Data& item) {
       process(item);
                               移植性に優れる
     });
  } catch (...) {
     // 並列throwされた例外のうち最初の1つ
対応コンパイラ: GCC, Clang, MSVC, ICC...
対応OS: Linux, Windows, MacOS, (Andorid), (iOS)
```

Microsoft ConcRT/PPL

```
#include <ppl.h>
 try {
   concurrency::parallel_for_each(begin(data), end(data),
   [](Data& item) {
      process(item);
   });
 } catch (...) {
   // 並列throwされた例外のうち最初の1つ
                         msvcrtにビルトイン
対応コンパイラ: MSVC
```

対応OS: Windows

35

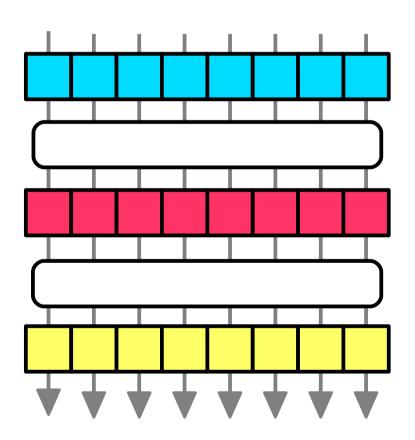
Intel CilkPlus

```
#include <cilk.h>
try {
  cilk_for (int i = 0; i < N; ++i) {
     process(data[i]);
                            記述性に優れる
} catch (...) {
  // 並列throwされた例外のうち最初の1つ
```

対応コンパイラ: ICC, (GCC), (Clang)

ベクトルレベル並列

(SIMD, GPGPU, etc.)



ベクトルレベル並列を支える技術

```
for (int i = 0; i < N; ++i) {
    A[i] += B[i] * C[i];
}
```

ライブラリ (+コンパイラサポート)

SIMD intrinsic, Boost.SIMD, Thrust, Bolt

言語拡張

CilkPlus, C++AMP, OpenMP 4.0, OpenACC

SIMD intrinsic (組込み関数)

```
#include <xmmintrin.h> // x86/SSE
vector<float> A, B, C;
int i, n = N \& ~3;
for (i = 0; i < n; i += 4) { // 4要素単位で処理
     m128 a4 = mm_loadu_ps(&A[i]);
     _{m128} b4 = _{mm_loadu_ps(\&B[i]);}
     m128 c4 = mm_loadu_ps(&C[i]);
   a4 = _{mm}add_{ps}(a4, _{mm}_{mul}_{ps}(b4, c4));
   _mm_storeu_ps(&A[i], a4);
for (; i < A.size(); ++i) // 端数処理
   A[i] += B[i] * C[i];
```

要 アセンブラ知識(どんな命令がある?)

命令セット毎に 異なるコードが必要 (SSE2, AVX, NEON...)

対応コンパイラ: GCC, Clang, MSVC, ICC...

Boost.SIMD (NT²)

```
#include <boost/simd/***.hpp>
vector<float, boost::simd::allocator<float>> A, B, C;
typedef boost::simd::pack<float> FVec;
int step = FVec::static_size;
int i, n = (N / step) * step;
for (i = 0; i < n; i += step) { // 処理単位は命令セット依存
   FVec a = boost::simd::aligned_load<FVec>(&A[i]);
   FVec b = boost::simd::aligned_load<FVec>(&B[i]);
   FVec c = boost::simd::aligned_load<FVec>(&C[i]);
   a += b * c;
                                          SIMD処理を一般化
   boost::simd::aligned_store(a, &A[i]);
                                          (x86/SSE系列,AVXと
for (; i < A.size(); ++i) // 端数処理
                                           PPC/AltiVec(こ対応)
   A[i] += B[i] * C[i];
```

ARM/NEON未対応

対応コンパイラ: GCC, Clang, MSVC, ICC

Intel CilkPlus / Array Notation

Array Notation未対応

対応コンパイラ: ICC, (GCC), (Clang)

Thrust (GPGPU; CUDA)

```
CPU向け(TBB, OpenMP)
 #include <thrust/***.h>
                                          バックエンドも提供
thrust::device_vector<float> A, B, C;
struct fma_functor {
    template <typename Tuple>
    __device__ void operator()(Tuple t) {
       thrust::get<0>(t) += thrust::get<1>(t) * thrust::get<2>(t);
    } // Tuple = [A, B, C]
};
thrust::transform(
    thrust::make_zip_iterator(thrust::make_tuple(begin(A), begin(B), begin(C))),
    thrust::make_zip_iterator(thrust::make_tuple(end(A), end(B), end(C))),
    fma functor());
対応コンパイラ: CUDA Compiler, GCC, Clang, MSVC
```

Microsoft C++ AMP (GPGPU; DirectX)

```
#include <amp.h>
vector<float> A, B, C;
concurrency::array_view<float, 1> a(N, &A[0]);
concurrency::array_view<float, 1> b(N, &B[0]);
concurrency::array_view<float, 1> c(N, &C[0]);
concurrency::parallel_for_each(
   a.extent,
   [=](concurrency::index<1> idx) restrict(amp) {
      a[idx] += b[idx] * c[idx];
                                          CPU/GPU処理を
                                           統一的に記述
);
a.synchronize(); // GPU→CPUデータ同期
```

対応コンパイラ: MSVC, (Clang)

スレッド/ベクトル 並列化技術

ライブラリ 言語拡張 **CilkPlus TBB** スレッド **PPL OpenMP** レベル並列 libstdc++ ParallelMode C++AMP Boost.SIMD ベクトル **CilkPlus Thrust** レベル並列 SIMD intrinsic

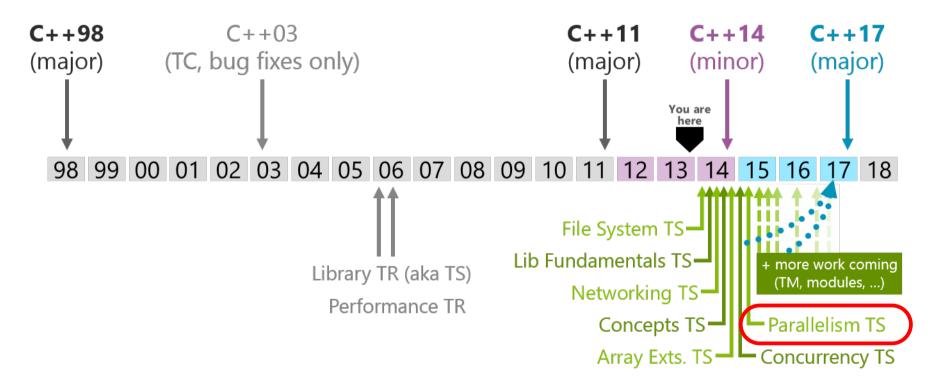
スレッドレベル並列 vs.

ベクトルレベル並列

スレッドレベル並列 ベクトルレベル並列 C++1y ParallelTS

C++並列拡張

N3850 "Working Draft, Technical Specification for C++ Extensions for Parallelism"



http://isocpp.org/blog/2013/10/trip-report-fall-iso-c-meeting

C++並列拡張

N3850 "Working Draft, Technical Specification for C++ Extensions for Parallelism"

並列アルゴリズムのインターフェースを規定

C++標準アルゴリズムを並列実行可能に**拡張** (<algorithm>, <numeric>, <memory>ヘッダ) API設計は **Thrust** をベースとする

C++並列拡張

N3850 "Working Draft, Technical Specification for C++ Extensions for Parallelism"

C++並列ライブラリの広範な実装経験確立が目的 検討用実装: https://github.com/n3554/n3554/

今は std::experimental::parallel 名前空間

→ std 名前空間 への昇格を目指す

#include <algorithm>

```
using namespace std;
try {
   for_each(
      begin(data), end(data), [](Data& item) {
         process(item);
      });
} catch (...) {
   // ...
```

```
#include <experimental/algorithm>
#include <experimental/execution_policy>
using namespace std::experimental::parallel;
try {
   for_each(
     par, // 並列実行ポリシー
     begin(data), end(data), [](Data& item) {
        process(item);
     });
} catch (exception_list& exlist) {
  for (exception_ptr ex : exlist) {
     // 並列throwされた全ての例外が捕捉される
```

```
// C++標準アルゴリズム
template <class In, class F>
In for_each( In first, In last, F f );
```

実行ポリシーを指定する オーバーロード追加

// C++1y ParallelTS template <class ExecPolicy, class In, class F> In for_each(ExecPolicy&& exec, In first, In last, F f);

<algorithm></algorithm>					
all_of					
none_of					
for_each					
for_each_n					
find					
find_if					
find_if_not					
find_end					
find_first_of					
adjacent_find					
count					
count_if					
mismatch					
equal					
search					
search_n					
сору					
copy_n					
move					

swap range

transform replace replace_if replace_copy replace_copy_if fill fill n generate generate_n remove remove if remove_copy remove_copy_if unique unique copy reverse reverse copy rotate rotate_copy is_partitioned partition

stable_partition partition copy partition_point sort stable sort partial_sort partial_sort_copy is sorted is sorted until nth_element merge inplace_merge includes set union set intersection set difference set_symmetric_difference min_element max element minmax_element

lexicographical compare

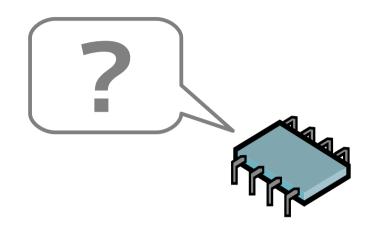
<numeric> reduce inner_product exclusive scan inclusive scan adjacent_difference <memory> uninitialized_copy uninitialized_copy_n uninitialized fill uninitialized fill n

一部は明示的に除外 <algorithm> shuffle <numeric> accumulate etc.

実行ポリシー

seq 逐次実行

呼び出しスレッド上で実行



par 並列実行

複数スレッド上で順序付けされない実行(*unordered*) もしくは 1スレッド上で非決定順の実行(*indeterminately sequenced*)

vec ベクトル実行

複数スレッド上で順序付けされない実行(*unordered*) もしくは 1スレッド上で序列化されない実行(*unsequenced*)

実行ポリシー

seq 逐次実行

シングルスレッド実行(従来C++標準アルゴリズムと同一)

par 並列実行

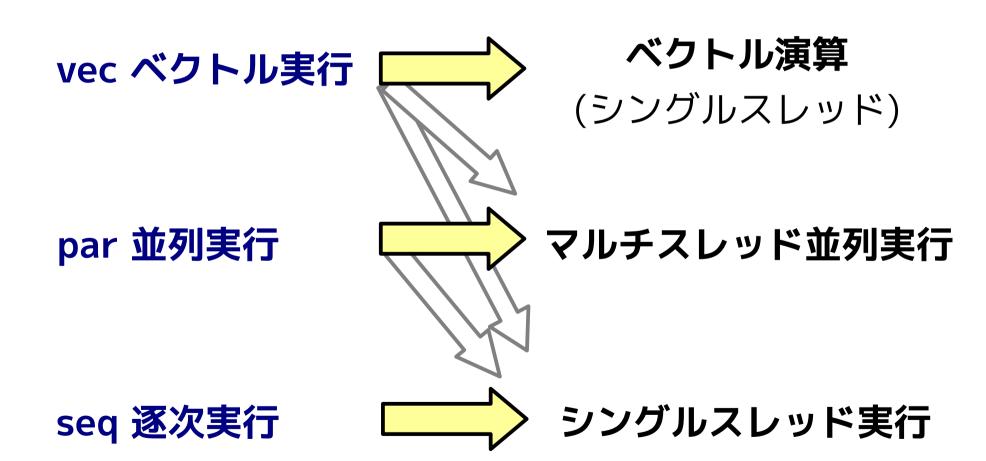
マルチスレッドによる並列実行 または シングルスレッド実行(実行順序は同不順)

vec ベクトル実行

1スレッド上で**ベクトル並列**実行 または **マルチスレッド**による並列実行 または **シングルスレッド**実行(実行順序は同不順)

実行ポリシーと実行形態

コンパイラやライブラリが実行形態を選ぶ



実行ポリシー / static

コンパイル時に実行ポリシーを指定

```
vector<float> data = ...;
// 逐次ソート
sort( seq, begin(data), end(data) );
// sort( begin(data), end(data) ); と等価
// マルチスレッドによる並列ソート
sort( par, begin(data), end(data));
// ベクトル演算器によるソート
sort( vec, begin(data), end(data) );
```

実行ポリシー / dynamic

ランタイムに実行ポリシーを選択

```
vector<float> data = ...;

// デフォルトは逐次ソート

execution_policy exec = seq;

if (data.size() > threshold) {

// データ数が多ければ並列ソートに切替

exec = par;

}

sort( exec, begin(data), end(data));
```

```
#include <experimental/algorithm>
#include <experimental/execution_policy>
using namespace std::experimental::parallel;
try {
  for_each(
     par, // 並列実行ポリシー
     begin(data), end(data), [](Data& item) {
        process(item);
     });
} catch (exception_list& exlist) {
  for (exception_ptr ex : exlist) {
     // 並列throwされた全ての例外が捕捉される
```

exception_list

並列アルゴリズムからthrowされうる例外型 並列にthrowされた全ての例外オブジェクトを格納

```
class exception_list: public exception
public:
   size t size() const;
   iterator begin() const;
   iterator end() const;
private:
   list<exception_ptr> exceptions_; // exposition only
};
```

実行ポリシーとC++例外

seq 逐次実行 と par 並列実行

関数オブジェクトからの例外は、exception_list型に補足されて呼び出し元へthrowされる。

vec ベクトル実行

関数オブジェクトから例外throwされた場合、std::terminate

が呼び出されてプログラム終了する。



実行ポリシーとC++例外

seq 逐次実行 と par 並列実行

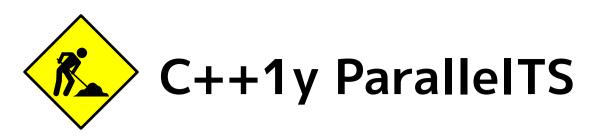
任意のスレッド上で実行されうるセマンティクス 複数スレッドからthrowされた例外を全て補足

vec ベクトル実行

ベクトル並列化(SIMD, GPGPUなど)可能なセマンティクスコンパイラによるベクトル化を阻害しないよう例外使用禁止

実行ポリシーと制約事項

	並列性の セマンティクス	処理 順序	例外 throw	関数オブジェクト 制約事項
vec ベクトル実行	ベクトル演算 (<i>シングルスレッド</i> <i>上の並列処理</i>)	順不同	×	相互干渉は全てNG
par 並列実行	マルチスレッド 並列処理	順不同		順序依存性はNG (<i>Ex. 他反復の待機処理</i> でDeadLock発生)
seq 逐次実行	シングルスレッド 逐次処理	安定	0	特になし標準アルゴリズムと同じ



TS(Technical Specification)はC++標準ではありませんしかも**ParalleITS**はまだWorking Draft段階です

→ 今後もいろいろ仕様変更されます

参考までに:

C++ Transactional Memory拡張(@Boost.勉強会#10)

http://www.slideshare.net/yohhoy/boostjp10-tm-20120728

- → 紹介当時(2012/7)と現在で大きく様変わりしましたホントに
- → 基本コンセプトの名称変更・セマンティクス修正、 キーワード・属性の全面変更、etc.

Need More Speed?

速度性能はコンパイラやライブラリ実装に依存します

ただし、一般論として…

CPU bound (I/O boundは並行処理向け)

並列化可能なアルゴリズムへの変更

タスク間の依存性低減(同期箇所・共有データの削減)

データメモリ配置(局所性、偽共有、アライメント)
Locality False sharing Alignment

計測!計測!計測!

まとめ

宣言的なコード記述による**並列ループ**構文

スレッドレベル並列の技術

TBB, PPL, CilkPlus, OpenMP, etc.

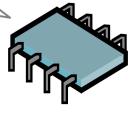
ベクトルレベル並列の技術

Boost.SIMD, Thrust, CilkPlus, C++AMP,

OpenMP 4.0, OpenACC, Bolt, etc.

ParallelTSではスレッド/ベクトル並列統合を目指す?

ライブラリ実装とコンパイラ技術 今後の発展にご期待ください



参考URL(1)

n3850 "Working Draft, Technical Specification for C++ Extensions for Parallelism" (Thrustベース 並列アルゴリズム提案)

http://www.open-std.org/jtc1/sc22/wg21/docs/papers/2014/n3850.pdf https://github.com/n3554/n3554/

n3571 "A Proposal to add Single Instruction Multiple Data Computation to the Standard Library" (Boost.SIMDベース データ型+操作関数提案)

http://www.open-std.org/jtc1/sc22/wg21/docs/papers/2013/n3571.pdf

n3831 "Language Extensions for Vector level parallelism"

(CilkPlusベース ベクトル並列ループ構文提案)

http://www.open-std.org/jtc1/sc22/wg21/docs/papers/2014/n3831.pdf

n3851 "Multidimensional bounds, index and array_view"

(C++AMPベース データ型提案)

http://www.open-std.org/jtc1/sc22/wg21/docs/papers/2014/n3851.pdf

参考URL(2)

OpenMP http://openmp.org/

Thrust http://thrust.github.io/

Intel TBB(Threading Building Blocks)

https://www.threadingbuildingblocks.org/

Microsoft ConcRT(Concurrency Runtime) / PPL(Parallel Pattern Library)

http://msdn.microsoft.com/ja-jp/library/dd492418.aspx

libstdc++ Parallel Mode

http://gcc.gnu.org/onlinedocs/libstdc++/manual/parallel_mode.html

Intel CilkPlus https://www.cilkplus.org/

GCC実装 https://www.cilkplus.org/build-gcc-cilkplus

Clang実装 http://cilkplus.github.io/

Microsoft C++ AMP(Accelerated Massive Parallelism)

http://msdn.microsoft.com/ja-jp/library/hh265137.aspx

Clang http://isocpp.org/blog/2012/12/shevlin-park

Boost.SIMD https://github.com/MetaScale/nt2/

SIMD intrinsic function

x86 http://software.intel.com/sites/landingpage/IntrinsicsGuide/

ARM http://infocenter.arm.com/help/topic/com.arm.doc.dui0491c/CIHJBEFE.html

[予備] OpenMPでC++例外伝播

```
#include <omp.h>
try {
    std::exception ptr ep;
    omp_lock_t ep_guard;
    omp_init_lock(&ep_guard);
    #pragma omp parallel for
    for (int i = 0; i < N; ++i) {
        try {
             process(data[i]); // C++例外をthrow
        } catch (...) {
             omp_set_lock(&ep_guard);
             if (!ep) ep = std::current_exception(); // 最初の例外のみ保存
             omp_unset_lock(&ep_guard);
    }
    omp_destroy_lock(&ep_guard);
    if (ep) std::rethrow_exception(ep); // 例外をparallel region外で再throw
} catch (...) {
    // 例外処理
```

[予備] OpenMP vs. CilkPlus/TBB/PPL

CilkPlus, TBB, PPL

スレッドプール+Work Stealing 実行時に論理タスク→物理スレッドへのマッピング 高いスケーラビリティを得やすい

OpenMP

parallel指示文で明示的にスレッド生成(≠宣言的) スレッドプールによる実装も存在(処理系の品質) OpenMP 3.0で"task"導入(※MSVCは非対応)

[予備] OpenMP 4.0, n3831

```
// 逐次処理
for (int i = 0; i < N; ++i) {
   A[i] += B[i] * C[i];
                      // OpenMP 4.0 / SIMD construct
                      #pragma omp simd aligned(A,B,C:16)
                      for (int i = 0; i < N; ++i) {
                         A[i] += B[i] * C[i];
                      }
                      // n3831 Language Extensions
                      // for Vector level parallelism
                      for simd (int i = 0; i < N; ++i) {
                         A[i] += B[i] * C[i];
```