並列有限要素法による 三次元定常熱伝導解析プログラム 並列可視化

中島 研吾 東京大学情報基盤センター

自動チューニング機構を有する アプリケーション開発・実行環境 ppOpen-HPC

中島研吾

東京大学情報基盤センター

佐藤正樹(東大・大気海洋研究所),奥田洋司(東大・新領域創成科学研究科), 古村孝志(東大・情報学環/地震研),岩下武史(北大・情報基盤センター), 阪口秀(海洋研究開発機構)

背 景(1/2)

- 大規模化、複雑化、多様化するハイエンド計算機
 環境の能力を充分に引き出し、効率的なアプリ
 ケーションプログラムを開発することは困難
- 有限要素法等の科学技術計算手法:
 - プリ・ポスト処理, 行列生成, 線形方程式求解等の一 連の共通プロセスから構成される。
 - これら共通プロセスを抽出し、ハードウェアに応じた最 適化を施したライブラリとして整備することで、アプリ ケーション開発者から共通プロセスに関わるプログラ ミング作業、並列化も含むチューニング作業を隠蔽で きる。
 - -アプリケーションMW, HPC-MW, フレームワーク

背 景(2/2)

- A.D.2000年前後
 - GeoFEM, HPC-MW
 - 地球シミュレータ, Flat MPI, FEM
- ・現在:より多様,複雑な環境
 - ーマルチコア, GPU
 - ハイブリッド並列
 - MPIまでは何とかたどり着いたが・・・
 - ・「京」でも重要
 - CUDA, OpenCL, OpenACC
 - ポストペタスケールからエクサス ケールへ
 - ・より一層の複雑化





HPCミドルウェア:何がうれしいか

- アプリケーション開発者のチュー
 ニング(並列,単体)からの解放
 - <u>SM</u>ASHの探求に専念
 - 一生SMASHと付き合うのはきつい
 - SM<u>ASH</u>をカバー
- コーディングの量が減る
- 教育にも適している
- 問題点

- ハードウェア,環境が変わるたびに 最適化が必要となる



Hardware

ppOpen-HPC

- ppOpen-HPCはJST戦略的創造研究推進事業CREST研究領域 「ポストペタスケール高性能計算に資するシステムソフトウェア技術 の創出(研究総括:佐藤三久(理研AICS))」の一環として「自動 チューニング機構を有するアプリケーション開発・実行環境」におい て2011年度から5年計画で開発されているオープンソースフレーム ワーク(pp=post peta)であり、
- メニィコアアーキテクチャに基づくポストペタスケールシステムの処理 能力を充分に引き出す科学技術アプリケーションの効率的な開発, 安定な実行に資するものである。
- 対象離散化手法を有限要素法等の5手法に限定し、各手法の特性に基くアプリケーション開発用ライブラリ群、実行環境を提供する。
- 本研究は2016年度に筑波大学,東京大学共同で導入予定のポスト T2Kシステム(ピーク性能30PFLOPS以上)をターゲットとし,東大情 報基盤センタースパコンの2,000人以上の利用者の新システムへの 円滑な移行を支援する。





ppOpen-HPC covers ...





ppOpen-HPC

- ppOpen-HPCはJST戦略的創造研究推進事業CREST研究領域 「ポストペタスケール高性能計算に資するシステムソフトウェア技術 の創出(研究総括:佐藤三久(理研AICS))」の一環として「自動 チューニング機構を有するアプリケーション開発・実行環境」におい て2011年度から5年計画で開発されているオープンソースフレーム ワークであり.
- メニィコアアーキテクチャに基づくポストペタスケールシステムの処理 能力を充分に引き出す科学技術アプリケーションの効率的な開発. 安定な実行に資するものである。
- 対象離散化手法を有限要素法等の5手法に限定し、各手法の特性 に基くアプリケーション開発用ライブラリ群、実行環境を提供する。
- 本研究は2016年度に筑波大学,東京大学共同で導入予定のポスト T2Kシステム(ピーク性能30PFLOPS以上)をターゲットとし、東大情 報基盤センタースパコンの2,000人以上の利用者の新システムへの 円滑な移行を支援する。





ppOpen-HPCのターゲット ポストT2K

• ポストT2K

University of Tsukuba

- ピーク性能30 PFLOPS超, (多分)2016年秋稼働開始
 - ✓ 最先端共同HPC基盤施設(JCAHPC, Joint Center for Advanced High Performance Computing): 筑波大学,東京大学

✓ <u>http://jcahpc.jp/</u>

- メニーコアベース (Intel MIC/Xeon Phi等)

✓ MPI + OpenMP + X



- <u>ppOpen-HPCは利用者(東大センター:2,000人)の新システムの</u> <u>円滑な移行を促進する</u>
- 京/FX10, Cray, Xeonクラスターも視野



ポストT2K: Oakforest-PACS

http://www.cc.u-tokyo.ac.jp/system/ofp/

- 2016年12月1日稼働開始
- ・ 8,208 Intel Xeon/Phi (KNL), ピーク性能25PFLOPS
 - 富士通が構築
- <u>最先端共同HPC 基盤施設(JCAHPC: Joint Center</u> for Advanced High Performance Computing)
 - 筑波大学計算科学研究センター
 - 東京大学情報基盤センター
 - 東京大学柏キャンパスの東京大学情報基盤センター内に、両機関の 教職員が中心となって設計するスーパーコンピュータシステムを設置
 し、最先端の大規模高性能計算基盤を構築・運営するための組織
 - <u>http://jcahpc.jp</u>





ppOpen-HPCでは北大・東大・京大の各大型計算機センターと、地球科学(東大(大気海洋研、地震研)、JAMSTEC)、工学(東大(新領域))分野の密接な協力のもとで、実用的な大規模アプリケーション開発フレームワーク構築を目指している。

様々な分野の専門家によるCo-Design

- 地球シミュレータ, 京コンピュータ, T2K等で既に稼働実績のある大 規模シミュレーションコードの共通機能を取り出し, 次世代システム 向けの最適化, 自動チューニング機構の適用を実施している。
- 開発成果はソースレベルで公開すると共に、東大情報基盤センター、 PCクラスタコンソーシアム(PCクラスタ実用アプリケーション部会)共催により、産業利用も目指した講習会を開催し普及を図っている。



- FEM等の科学技術計算手法 はいくつかの典型的な計算プ ロセスを含む
 - 各プロセスの最適化は可能で あり、アプリ開発において有用
- データ構造を適切に設計すれば、並列計算における通信の「隠蔽」が可能
 - 並列有限要素法におけるHalo
- GeoFEM, HPC-MW等のフレームワークによってアプリ開発者を最適化,並列化から解放することが可能
- チューニングの継承が問題





ppOpen-HPC開発方針

- 先行研究において各メンバーが開発した大規模アプリ ケーションに基づきppOpen-APPLの各機能を開発,実装
 - 各離散化手法の特性に基づき開発・最適化
 - ・共通データ入出カインタフェース、領域間通信、係数マトリクス生成
 - ・離散化手法の特性を考慮した前処理付き反復法
 - 適応格子, 動的負荷分散
 - 実際に動いているアプリケーションから機能を切り出す
 - 各メンバー開発による既存ソフトウェア資産の効率的利用
 - GeoFEM, HEC-MW, HPC-MW, DEMIGLACE, ABCLibScript
- ppOpen-ATはppOpen-APPLの原型コードを対象として 研究開発を実施し、その知見を各ppOpen-APPLの開発、 最適化に適用
 - 自動チューニング技術により,様々な環境下における最適化ライ ブラリ・アプリケーション自動生成を目指す 15



・5種類の離散化手法に対応

- 連成計算も可能

- 自動チューニング(AT)の採用
 - チューニング技術の継承
 - アーキテクチャの(ものすごく大きくない)変化に自動的に対応 - 問題サイズ等の細かい設定に最適なコードを生成できる



- 本プロジェクトの目的はアプリ開発ではないが、検証用、潜 在ユーザー向けにターゲットアプリを設定
- ppOpen-APPL/FEM
 - 非圧縮性流体, 熱伝導(定常, 非定常), 固体力学(動的, 静的)
- ppOpen-APPL/FDM
 - 非圧縮性流体, 熱伝導(非定常), 固体力学(動的)
- ppOpen-APPL/FVM
 - 圧縮性流体, 熱伝導(定常)
- ppOpen-APPL/BEM
 - 電磁気学, 固体力学(準静的): 地震発生サイクル
- ppOpen-APPL/DEM
 - 非圧縮性流体, 固体力学(動的)



18

Level-0



コード公開スケジュール (英語ドキュメント付き, MITライセンス) http://ppopenhpc.cc.u-tokyo.ac.jp/

- 毎年のSC-XYで更新,公開
- Flat MPI, OpenMP/MPIハイブリッド並列
- Multicore/Manycoreクラスタ向け→Xeon Phi最適化

公開の履歴

- SC12, Nov 2012 (Ver.0.1.0)
- SC13, Nov 2013 (Ver.0.2.0)
- SC14, Nov 2014 (Ver.0.3.0)
- SC15, Nov 2015 (Ver.1.0.0)



New Features in Ver.1.0.0

http://ppopenhpc.cc.u-tokyo.ac.jp/

- HACApK library for H-matrix comp. in ppOpen-APPL/BEM (OpenMP/MPI Hybrid Version)

 First Open Source Library by OpenMP/MPI Hybrid
- ppOpen-MATH/MP (Coupler for Multiphysics Simulations, Loose Coupling of FEM & FDM)
- Matrix Assembly and Linear Solvers for ppOpen-APPL/FVM





研究協力·普及

- 国際的共同研究
 - Lawrence Berkeley National Lab.
 - 国立台湾大学, 国立中央大学(台湾)
 - ESSEX/SPPEXA/DFG, Germany
 - IPCC (Intel Parallel Computing Ctr.)
- 普及
 - 大規模シミュレーションへの適用
 - CO₂ 地下貯留, 物性物理
 - ・ 宇宙物理, 地震シミュレーション
 - ppOpen-AT, ppOpen-MATH/VIS, ppOpen-MATH/MP, 線形ソルバー群
 - H行列ライブラリ
 - 国際WS(2012,13,15)
 - 講習会(東大センター), 講義







共同研究等事例(1/2)

- ppOpen-AT関連共同研究
 - 工学院大学 田中研究室
 - 田中研究室開発のAT方式(d-spline方式)の適用対象としてppOpen-ATのAT機能を拡張
 - 東京大学 須田研究室
 - ・電力最適化のため、須田研究室で開発中のAT方式と電力測定の共通 APIを利用し、ppOpen-ATを用いた電力最適化方式を提案[Katagiri et al. IEEE/MCSoC 2013 Best Paper Award]
- JHPCN共同研究課題
 - 高精度行列 行列積アルゴリズムにおける並列化手法の開発 (東大, 早稲田大)(H24年度)(研究としては継続)
 - ・高精度行列-行列積演算における行列-行列積の実装方式選択に利用
 - 粉体解析アルゴリズムの並列化に関する研究(東大,法政大) (H25年度)
 - 粉体シミュレーションのための高速化手法で現れる性能パラメタのATで 利用を検討



共同研究等事例(2/2)

- JHPCN共同研究課題(続き)
 - 巨大地震発生サイクルシミュレーションの高度化(京大,東大他)(H24・25年度)
 - Hマトリクス, 領域細分化
 - ポストペタスケールシステムを目指した二酸化炭素地中貯留シ ミュレーション技術の研究開発(大成建設,東大)(H25年度)
 - 疎行列ソルバー, 並列可視化
 - 降着円盤シミュレーション(千葉大,東大)(H22年度~)
 - 疎行列ソルバー, 並列可視化
 - 本陽磁気活動の大規模シミュレーション」(東大(地球惑星,情報 基盤センター))(H25年度~)
 - 疎行列ソルバー, 並列可視化





ppOpen-MATH

- ppOpen-HPCにおける共通ライブラリ
- ppOpen-MATH/MG
 - 多重格子法ソルバー
- ppOpen-MATH/GRAPH
 - 並列グラフライブラリ(マルチスレッド版):RCM, MC(開発中)
- ppOpen-MATH/VIS
 - 並列可視化ライブラリ
- ppOpen-MATH/MP
 - 弱連成カプラー

ppOpen-HPCにおける 並列可視化の考え方



ppOpen-HPCにおける 並列可視化の考え方



並列可視化とスパコン

- ・「見る」ためにスパコンは使わない
- ・「絵を出すために計算をやり直す」という考え方も採らない
- 大型計算機センターとしては、つぎ込めるだけの予算を計 算エンジンにつぎ込みたい

ppOpen-MATH/VIS

- ボクセル型背景格子を使用した大規模並列可視化手法 [Nakajima & Chen 2006]に基づく
 - 差分格子用バージョン公開:ppOpen-MATH/VIS-FDM3D
- UCD single file
- プラットフォーム
 - T2K, Cray
 - FX10
 - Flat MPI
 - Hybrid, 非構造格子:今年度実施



[Refine]

- AvailableMemory = 2.0 MaxVoxelCount = 500 MaxRefineLevel = 20
 - Available memory size (GB), not available in this version.
 - 0 Maximum number of voxels
 - = 20 Maximum number of refinement levels

Simplified Parallel Visualization using Background Voxels

- Octree-based AMR
- AMR applied to the region where gradient of field values are large
 - stress concentration, shock wave, separation etc.
- If the number of voxels are controled, a single file with 10⁵ meshes is possible, even though entire problem size is 10⁹ with distributed data sets.







Voxel Mesh (adapted)



Flow around a sphere



Example of Surface Simplification



FEM Mesh (SW Japan Model)



pFEM3D + ppOpen-MATH/VIS

<u>FORTRANユーザー</u>

- >\$ cd ~/pFEM/srcV
- >\$ make
- >\$ cd ../run
- >\$ pjsub gv.sh

<u> cユーザー</u>

- >\$ cd ~/pFEM/srcV
- >\$ make
- >\$ cd ../run
- >\$ pjsub gv.sh

Makefile(Fortran)

```
include Makefile.in
FFLAGSL = -I/home/S11502/nakajima/pfem/include
FLDFLAGSL = -L/home/S11502/nakajima/pfem/lib
LIBSL = -lfppohvispfem3d -lppohvispfem3d
.SUFFIXES:
.SUFFIXES: .o .f90 .f
.f.o:
        (FC) -c (FFLAGS) (FFLAGSL) < -o @
.f90.o:
        $(F90) -c $(F90FLAGS) $(FFLAGSL) $< -o $@
TARGET = .../run/solv
OBJS = Y
       test1.o ...
all: $(TARGET)
$(TARGET): $(OBJS)
        (F90) - 0  (TARGET)  (F90FLAGS)  (FFLAGSL)  (OBJS)  (LDFLAGSL)
$(LIBS) $(LIBSL) $(FLDFLAGSL)¥
```

Makefile(C)

```
include Makefile.in
CFLAGSL = -I/home/S11502/nakajima/pfem/include
LDFLAGSL = -L/home/S11502/nakajima/pfem/lib
LIBSL = -lppohvispfem3d
.SUFFIXES:
.SUFFIXES: .o .c
.C.O:
        $(CC) -c $(CFLAGS) $(CFLAGSL) $< -o $@
TARGET = .../run/solv
OBJS = Y
        test1.o ...
all: $(TARGET)
$(TARGET): $(OBJS)
        $(CC) -o $(TARGET) $(CFLAGS) $(CFLAGSL) $(OBJS) $(LDFLAGSL)
$(LIBS) $(LIBSL)
        rm - f *.o *.mod
```

pFEM-VIS

Fortran/main (1/2)

```
use solver11
use pfem_util
use ppohvis pfem3d util
 implicit REAL*8(A-H,O-Z)
 type(ppohVIS BASE stControl)
                                           :: pControl
 type(ppohVIS BASE stResultCollection)
                                           :: pNodeResult
 type(ppohVIS BASE stResultCollection)
                                           :: pElemResult
character(len=PPOHVIS BASE FILE NAME LEN) :: CtrlName
 character(len=PPOHVIS BASE FILE NAME LEN) :: VisName
character(len=PPOHVIS BASE LABEL LEN)
                                           :: ValLabel
                                            :: iErr
 integer(kind=4)
 CtrlName = ""
CtrlName = "vis.cnt"
VisName = ""
VisName = "vis"
ValLabel = ""
ValLabel = "temp"
call PFEM INIT
 call ppohVIS PFEM3D Init(MPI COMM WORLD, iErr)
 call ppohVIS PFEM3D GetControl(CtrlName, pControl, iErr);
 call INPUT CNTL
 call INPUT GRID
call ppohVIS PFEM3D SETMESHEX(
       NP,
                              NODE ID, XYZ,
&
                 N,
        ICELTOT, ICELTOT_INT, ELEM_ID, ICELNOD,
&
       NEIBPETOT, NEIBPE, IMPORT INDEX, IMPORT ITEM,
&
                           EXPORT INDEX, EXPORT ITEM, iErr)
&
```

&

&

&

&

pFEM-VIS

Fortran/main (2/2)

C/main (1/2)

```
#include <stdio.h>
#include <stdlib.h>
FILE* fp loq;
#define GLOBAL VALUE DEFINE
#include "pfem_util.h"
#include "ppohVIS PFEM3D Util.h"
extern void PFEM INIT(int,char**);
extern void INPUT CNTL();
extern void INPUT GRID();
extern void MAT CON0();
extern void MAT CON1();
extern void MAT ASS MAIN();
extern void MAT ASS BC();
extern void SOLVE11();
extern void OUTPUT UCD();
extern void PFEM FINALIZE();
int main(int argc, char* argv[])
  double START TIME, END TIME;
  struct ppohVIS FDM3D stControl *pControl = NULL;
  struct ppohVIS FDM3D stResultCollection *pNodeResult = NULL;
  PFEM_INIT(argc,argv);
  ppohVIS PFEM3D Init(MPI COMM WORLD);
  pControl = ppohVIS FDM3D GetControl("vis.cnt");
  INPUT CNTL();
  INPUT GRID();
  if(ppohVIS PFEM3D SetMeshEx(
      NP,N,NODE ID,XYZ,
      ICELTOT, ICELTOT INT, ELEM ID, ICELNOD,
      NEIBPETOT, NEIBPE, IMPORT INDEX, IMPORT ITEM, EXPORT INDEX, EXPORT ITEM)) {
                ppohVIS BASE PrintError(stderr);
                MPI Abort(MPI COMM WORLD,errno);
  };
```

C/main (2/2)

```
MAT CON0();
MAT CON1();
MAT ASS MAIN();
MAT_ASS_BC() ;
SOLVE11();
OUTPUT_UCD();
pNodeResult=ppohVIS BASE AllocateResultCollection();
      if(pNodeResult == NULL) {
              ppohVIS BASE PrintError(stderr);
              MPI Abort(MPI COMM WORLD,errno);
      };
if(ppohVIS_BASE_InitResultCollection(pNodeResult, 1)) {

              MPI Abort(MPI COMM WORLD,errno);
      };
      pNodeResult->Results[0] =
ppohVIS PFEM3D ConvResultNodeItemPart(NP,1,0,"temp",X);
START TIME= MPI Wtime();
      if(ppohVIS PFEM3D Visualize(pNodeResult,NULL,pControl,"vis",1)) {
              ppohVIS BASE PrintError(stderr);
              MPI Abort(MPI COMM WORLD,errno);
      };
ppohVIS PFEM3D Finalize();
PFEM_FINALIZE() ;
```

pFEM3D + ppOpen-MATH/VIS



分散メッシュファイルの準備

>\$ cd <\$O-TOP>/pfem3d/pmesh
(mesh.inp, mg.sh)

```
>$ pjsub mg.sh
```

<u>mesh.inp</u> 256 256 256 4 4 4

pcube

256³のメッシュを4×4×4=64分割 各MPIプロセスは64³

mg.sh

```
#!/bin/sh
#PJM -L "node=4"
#PJM -L "elapse=00:05:00"
#PJM -L "rscgrp=school"
#PJM -j
#PJM -j
#PJM -o "mg.lst"
#PJM --mpi "proc=64"
mpiexec ./pmesh
rm wk.*
```

計算の実行+可視化

```
>$ cd <$O-TOP>/pfem3d/run
(INPUT.DAT, gv.sh)
```

```
>$ pjsub gv.sh
```

INPUT.DAT

../pmesh/pcube
2000
1.0 1.0
1.0e-08

gv.sh

#!/bin/sh
#PJM -L "node=4"
#PJM -L "elapse=00:10:00"
#PJM -L "rscgrp=school"
#PJM -j
#PJM -j
#PJM -o "aa.lst"
#PJM --mpi "proc=64"

mpiexec ./solv

vis.cnt

[Refine]	
AvailableMemory	= 2.0
MaxVoxelCount =	1000
MaxRefineLevel	= 20
[Simple]	
ReductionRate =	0.0
[Output]	
FileFormat =	2

細分化制御情報セクション 利用可能メモリ容量(GB) not in use Max Voxel # Max Voxel Refinement Level 簡素化制御情報セクション 表面パッチ削減率 出力形式 =1:MicroAVS, =2:ParaView

