

2016.1.5 第4回CMSI人材育成シンポジウム

「高度計算科学技術を有する人材が活躍する社会を目指して ～国際競争力強化のためのICTの浸透～」

# 大学における計算機教育とHPC

---

藤堂眞治

東京大学大学院理学系研究科 / 物性研究所

wistaria@phys.s.u-tokyo.ac.jp



# どのような立場からの話か？

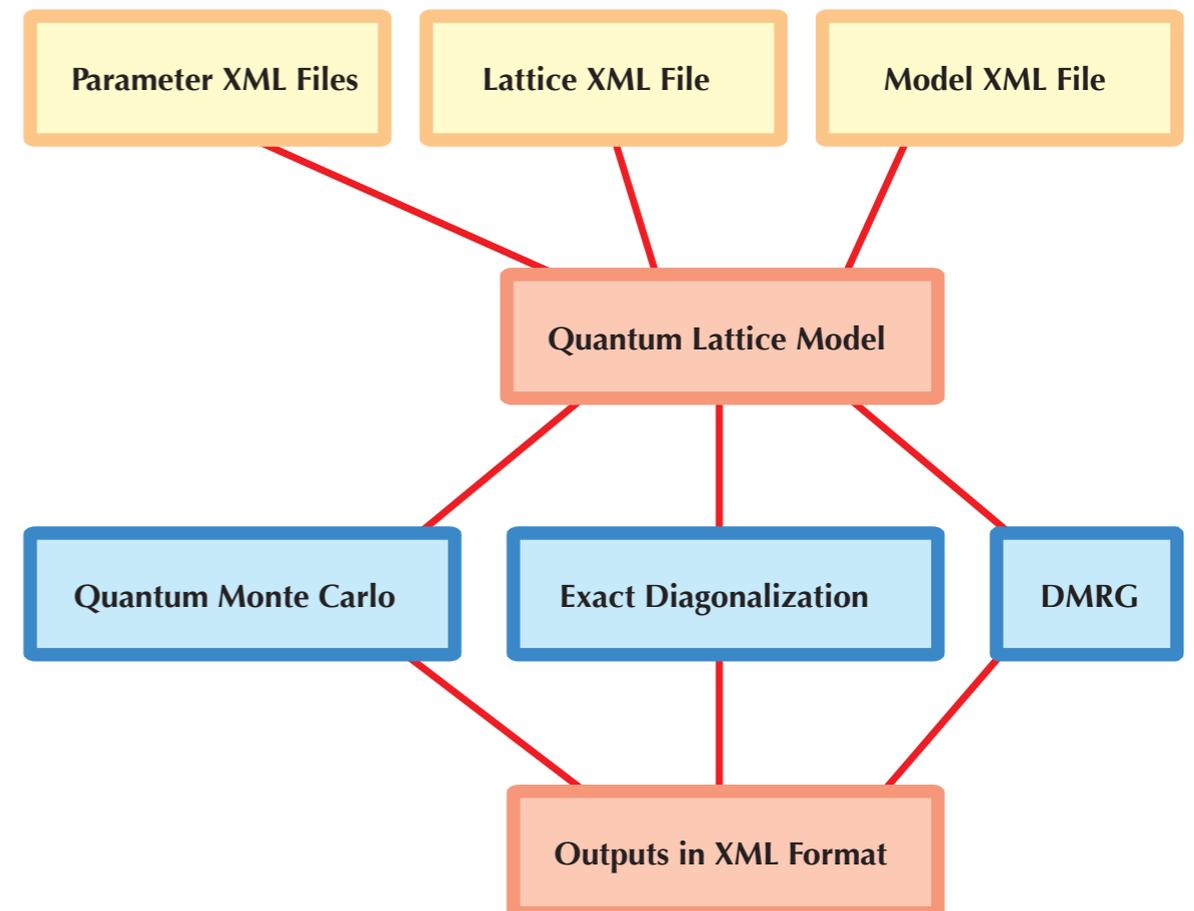
---

- 藤堂眞治(とうどうしんじ) — 東京大学大学院理学系研究科 / 物性研究所
  - CMSI重点課題「相関の強い量子系の新量子相探索とダイナミックスの解明」  
研究担当者
    - 大規模並列量子モンテカルロ法ALPS/looper開発
  - 量子多体系シミュレーションのためのオープンソースソフトウェアALPS開発者
  - CMSI広報小委員会代表・(元)神戸拠点代表
    - 人材育成・分野振興活動: アプリ講習会、若手技術交流会、MateriApps運営
  - 東京大学理学部物理学科教員
    - 学部講義「計算機実験」担当

# ALPS プロジェクト

ALPS = **A**lgorithms and **L**ibraries for **P**hysics **S**imulations

- 量子スピン系、電子系など強相関量子格子模型のシミュレーションのためのオープンソースソフトウェアの開発を目指す国際共同プロジェクト
- **ALPS ライブラリ** = C++による格子模型のための汎用ライブラリ群
- **ALPS アプリケーション** = 最新のアルゴリズムに基づくアプリケーション群: QMC、DMRG、ED、DMFT 等
- **ALPS フレームワーク** = 汎用入出力形式、解析ツール、スケジューラなど、大規模並列シミュレーションのための環境



# ALPS の生まれた背景

---

- 「**コミュニティーコード**」の不在
  - 個々の研究者が独自のコードを作成・使用
  - 扱うモデル・格子毎にコードを作成
  - コードの再利用がほとんどなされない
- **アルゴリズムが複雑化**
  - 様々な新しいアルゴリズムの開発
  - プログラム開発の長期化
- **理論家/実験家**による使い易く、信頼性の高いシミュレーション パッケージへの強い要望
- (non-portableな) 独自形式による**入力・出力**の弊害

# ターゲット・オーディエンス

---

- 実験家
  - 物質のモデリングにソフトウェアパッケージを利用
  - 実験結果とシミュレーション結果のフィッティングにより、相互作用定数などを決定
- 理論家
  - 理論的なアイデアのチェックに使いやすい整備されたコードを利用
  - 自前のコードのデバッグに
  - 新しいコード開発の基盤としての利用
- 計算機科学者、学生、...

# ALPS の特徴

---

- **任意の**格子
  - XML フォーマットによる格子構造の定義
  - ユニットセルの繰り返しによる格子生成
  - 格子を任意の有限グラフ (頂点と辺の集合) として定義することも可能
- **任意の**ハミルトニアン (模型)
  - XML フォーマットによる量子数、演算子の定義
  - 数式によるハミルトニアンの定義
- 様々な**最新の**解法 (アプリケーション): ED、CMC、QMC、DMRG、DMFT
- **全ての** ALPS アプリケーションに共通の入力形式
- **汎用的な**出力形式、Python による解析・グラフ作成ツール

# 主要技術

---

- XML、HDF5による入出力
  - 可搬性 (ポータビリティ)
  - 出力結果の変換が容易・内容が一目瞭然
- C++ ジェネリックプログラミング
  - 柔軟性・再利用性
  - 高品質なコードの作成
- C++ 標準ライブラリ、Boostライブラリ、サードパーティーライブラリの利用
  - 開発のスピードアップ
  - 様々な標準アルゴリズム
- MPI + OpenMP による並列化
- スクリプト言語(Python)によるポスト処理
- 国際共同作業のためのインフラストラクチャー
  - オープンソースのツールの利用

# 「公開ソフト」への長い道のり

---

- ソースコード
- ビルドシステム
  - テストスイート
    - チュートリアル
      - Webページ
        - ユーザサポート
          - ライセンス
            - アプリ名
              - ロゴ
                - ドキュメント
                  - 講習会、、、、

# 開発・普及に役立つツールや仕組みの利用

---

- GitHub、SourceForge、SlideShare、YouTube
- MateriApps <http://ma.cms-initiative.jp/>
  - MateriApps web
  - MateriApps LIVE!
  - MateriApps Installer
  - MateriApps Cloud (coming soon?)
- コミュニティによるサポート
  - CMSIハンズオン開催サポート
  - 物性研 ソフトウェア高度化プロジェクト

# MateriApps — 物質科学シミュレーションのポータルサイト

・公開ソフトウェア(アプリケーション)を核としたコミュニティ形成をめざして



- ・155の物質科学アプリケーションやツールを紹介(2015年9月現在)
- ・「**やりたいこと**」からアプリケーションを検索
  - ・検索タグ：「特徴」「対象」「手法・アルゴリズム」
- ・**開発者の声**を利用者に届ける
  - ・アプリ紹介、開発者ページ、アプリの魅力・将来性・応用性
- ・フォーラム(掲示板)を利用した**意見交換**
- ・講習会情報・**web講習会**・更新情報
- ・月間 8000 ページビューにまで成長

2013年5月公開

# MateriApps 掲載アプリケーション

- 155の物質科学アプリケーションやツールを紹介 (2015年9月現在)

## 密度汎関数法

AkaiKKR☆

OpenMX☆

xTAPP☆

ABINIT☆

...

(37)

## 量子化学

FMO☆

SMASH☆

GAMESS☆

DC☆

...

(19)

## 分子動力学

MODYLAS☆

Gromacs☆

ERmod☆

MDACP

...

(19)

## 格子模型

ALPS☆

DSQSS

BLOCK

DMRG++

...

(22)

## 連続体シミュレーション

ANSYS Multiphysics

Octa ...

(8)

## データ解析

CLUPAN☆

phonopy☆ (26)

## 可視化

fu☆

TAPIOCA☆ (28)

☆ MateriApps LIVE! 収録 (一部予定) アプリ

# MateriApps 活動の目的

---

- 開発者側の問題点
    - 有益なプログラムはもっと使われるべきだが、多くのソフトは研究室内にとどまって終わる
    - 公開・情報発信には手間がかかる
    - アプリ開発を成果として主張しにくい(指標がない)
  - 利用者側の問題点
    - どんなプログラムがあるのかよくわからない
    - インストール・使い方について知りたい
    - 開発者の活動(特に講習会情報)をもっと知りたい
- 両者をつなぐ役割を果たしたい

# アプリケーション普及にむけた三本柱

- アプリの情報発信
  - ポータルサイト MateriApps web
- スパコン上でのアプリ利用支援
  - 「京」や国内主要スパコンへのアプリのプレインストール MateriApps Installer
- 個人・研究室レベルでのアプリ利用の支援
  - MateriApps LIVE!
- インストールや入力ファイルの準備における「壁」を解消
- 計算科学の専門家だけではなく、実験家や企業内の利用、教育活動における活用へ

# MateriApps 活動を通じて感じたこと

---

- (第一原理計算に限ったとしても)いろいろなアプリを俯瞰的に見られる人材の欠除
- MateriAppsなどで組織的にサポートできることとできないことがある
- 公開ソフトとして体をなしていないソフトも多数
  - ユーザー視点の欠除 (ドキュメント、サポート)
  - そもそも「プロジェクト」「プロダクト」という感覚がない。学んだこともない

# 大学における計算機教育の現状 (1/2)

---

- 計算科学アライアンス
  - 東大内の(現時点では)非公式な組織
  - 工学系、理学系、情報理工、情報基盤センター、新領域、物性研、他から参加
  - とりまとめ 今田先生(東大院工)
  - 学内の計算機教育の現状と共通シラバス、共通して使える講義資料の取りまとめ作業に着手
    - 複数の部局でバラバラに教えているもの
      - UNIX、C言語、Fortran
      - 計算機の数値表現、Newton法、補完・加速、数値微分、常微分方程式、線形方程式(直接法・反復法)、偏微分方程式、固有値問題(密行列・疎行列)、乱数、モンテカルロ積分



# 大学における計算機教育の現状 (2/2)

---

- 情報基盤センターなど一部の部局が中心になって教えているもの
  - 並列計算(MPI, OpenMP)、チューニング、スパコン実習、他
- あまり系統的に教えられてないもの、全く教えられてないもの
  - 最適化問題、データ解析、プログラム開発手法、デバッグ手法、プロジェクト・ソースコード管理
- 問題点
  - 計算機科学、自然科学、情報科学がそれぞれ専門化・細分化
  - 各分野の教員が片手間に計算機教育
  - 並列計算、大規模ソフトウェア開発がほとんど教えられていない
  - シラバスが計算機科学・計算科学の現状に追いついていない
    - 最適化問題やグラフアルゴリズム、社会科学分野・企業での計算機活用、など

# 新しい計算機学部教育の試み

---

- 東京大学理学部物理学科 学部3年夏学期 「計算機実験」
- これまでの経緯
  - 「物理学実験 I」 の一つの枠として 「計算機実験」 があった
  - 週3日午後 x 2週、20人程度のグループ
  - 初回にTAが簡単な説明、それ以外は実習の時間、2週間後にレポート提出
  - 内容 : UNIX操作、C言語、LaTeX、常微分方程式
- 2015年度から
  - 計算機教育の強化についての強い要望 (特に実験系の先生方から)
  - 「計算機実験」 として独立した講義に (週1コマ x 半年)
  - 物理学科3年夏学期・必修
  - 座学(教員1名) + 実習(助教2名・TA2名)形式に

# 「計算機実験」

- 講義・実習形式
  - 全体を2グループに分けて講義・実習を隔週で行う
- 講義(L1～L8)
  - 実習で必要となる基礎的な事項・アルゴリズム + より高度な話題
- 実習(EX1～EX6): 情報基盤センター大演習室
  - 端末を使ったプログラミング実習  
準備練習 + 基本課題 + 応用課題(オプション)
- レポート (計3回)
- 3～4人のグループに分け、グループワーク
  - 最終週に成果発表会、レポート

|      | グループX | グループY |
|------|-------|-------|
| 第1週  | L1    | L1    |
| 第2週  | EX1   | L2    |
| 第3週  | L2    | EX1   |
| 第4週  | EX2   | L3    |
| 第5週  | L3    | EX2   |
| 第6週  | EX3   | L4    |
| 第7週  | L4    | EX3   |
| 第8週  | L5    | EX4   |
| 第9週  | EX4   | L5    |
| 第10週 | L6    | EX5   |
| 第11週 | EX5   | L6    |
| 第12週 | L7    | EX6   |
| 第13週 | EX6   | L7    |
| 第14週 | L8    | L8    |

|   | 講義(L)   | 実習(EX)  |
|---|---|---|
| 1 | 講義・実習の概要、数値誤差、数値微分、ニュートン法、代数方程式   | リモートログイン、ファイルコピー、UNIX操作、エディタ、コンパイル、C言語、フィボナッチ数列、数値微分、ニュートン法 [テント写像、代数方程式]                 |
| 2 | バージョン管理システム、常微分方程式の初期値問題、Euler法、Runge-Kutta法、陰解法、シンプレクティック積分法                               | バージョン管理システム、グラフ作成、LaTeX、摩擦のあるバネ問題、中点法・3次・4次Runge-Kutta [シンプレクティック積分法、硬い方程式]               |
| 3 | Poisson方程式の境界値問題、連立一次方程式、Gaussの消去法、LU分解、反復法、Jacobi法、C言語におけるベクトルと行列、LAPACK                   | Gaussの消去法、LU分解、LAPACK、C言語におけるポインタ、ピボット選択、境界条件の設定、Jacobi法 [Gauss-Seidel法・SOR法、LAPACKとMKL]  |
| 4 | Schrödinger方程式の解法、Numerov法、行列の対角化、Jacobi法、Givens変換、Hermite行列、べき乗法、Rayleigh-Ritzの方法、Lanczos法 | ハウスホルダー法、固有ベクトルの直交性、べき乗法、二重井戸ポテンシャルの固有状態、単位と無次元化 [Lanczos法、疎行列に対する方法、Frank行列]             |
| 5 | 特異値分解、連立方程式の最小二乗解、行列の低ランク近似、画像圧縮、線形回帰分析、Ridge回帰、カーネル法、Bayes推定                               | 特異値分解、最小二乗フィッティング、CとFortranのバインディング、行列の低ランク近似、基底関数の追加、LU分解と特異値分解 [最小二乗法における誤差の評価、画像圧縮、実験] |
| 6 | 最適化問題、罫い込み法、最急降下法、共役勾配法、Nelder-Mead法、乱択アルゴリズム、モンテカルロ積分、重点的サンプリング、擬似乱数                       | 最適化問題、罫い込み法、最急降下法、Nelder-Mead法、擬似乱数、モンテカルロ積分  |
| 7 | MCMC、離散最適化、スパコンと計算物理  |   |
| 8 | グループワーク発表会  |   |

# 「計算機実験」における試み

- 実習環境
  - 情報基盤センター端末 + UNIXサーバ + 仮想マシン配布(MateriApps LIVE!)
- 伝統的な線形計算に加え、現代的な項目も追加
  - バージョン管理システム、ライブラリの利用、特異値分解、最適化問題、ベイズ推定、スパコンと並列計算
- バージョン管理を身につけるために、レポート(LaTeX)もバージョン管理を義務付ける
- 座学と実習を半々に
- 4人程度のグループに分け、自分たちでテーマを選んで「問題解決」に取り組む。最終回に発表会を実施



|    |                                 |
|----|---------------------------------|
| X1 | シミュレーテッド・アニーリングを用いた二次元イジング模型の解析 |
| X2 | 硬い方程式とその解法                      |
| X3 | 二重振り子のRunge-Kutta法による解析とアニメーション |
| X4 | 乱数を用いたブラウン運動の観察                 |
| X5 | ロジスティック方程式の差分化とカオス              |
| X6 | シンプレクティック積分法による振り子の解析           |
| X7 | シンプレクティック積分法(仮題)                |
| X8 | テント写像によるカオスの生成                  |
| X9 | Lanczos法で固有値を求める                |
| Y1 | パーコレーションモデルを用いた伝染病の解析           |
| Y2 | 非線形フィッティングによる内部抵抗の推定            |
| Y3 | 液体のシミュレーション                     |
| Y4 | SVD分解による画像ファイルの圧縮               |
| Y5 | ルンゲクッタ法による三体問題の8の字解の導出          |
| Y6 | Gauss-Seidel法とSOR法              |
| Y7 | いろいろな分布に従う乱数の発生                 |
| Y8 | ホップフィールドネットワークによる連想記憶           |
| Y0 | マルコフ連鎖モンテカルロ法によるXY模型の解析         |

- 学生からのフィードバック (授業アンケート他)
  - 授業の準備状況や教員の熱意は平均より高く評価されている
  - 難易度：高、進度：速
  - 「例題のソースコードを読むのが大変だった。最初から自分で書く方がよい」
  - 「バージョン管理システムに興味があったが、一回の講義・実習で実際に使えるようになるとは思わなかった」
  - 「五月祭の発表で、スパコンを使った並列計算にチャレンジしてみたい」
- 追加で、秋に4年生以上むけのインフォーマルな講義を実施
  - 三日間で短縮版の講義 + 実習
  - 単位なし、他学科、他専攻の学生も受講可
  - (自主的に) 55名が受講

# おわりに

---

- HPCの裾野拡大のためには、長期的な視点にたった学部教育の充実が必要
  - HPCを仕事とする人だけでなく、政策立案者、意思決定者、あるいはその側に現代的な計算機教育を受けた人が必要
- 現状ではシラバスは学科によってバラバラ。ある程度の統一と全体像の整理が必要
- 現代的な観点からの計算機教育の再構築が必要
  - 最適化問題、グラフ探索、データ解析、統計処理
  - 社会科学分野での応用
  - 企業でのシミュレーション活用を視野にいれた教育
  - 計算機を使った問題解決の体験
  - プロジェクト管理・ソースコード管理は研究倫理の観点からも重要