

惑星科学・宇宙科学におけるシミュレーション研究・ソフトウェア開発と探索・観測研究

牧野淳一郎

CPS

2021/11/9 惑星科学と情報基盤シンポジウム

— 惑星探査 **and/or** 系外惑星観測を支える計算惑星科学・データ惑星科学・**AI**
惑星科学・... —

話の構成

1. 「シミュレーション研究」の過去・現在・未来
2. 宇宙物理における大規模シミュレーションソフトウェアの現状
3. 我々の対応
4. 「キャリアパス」問題
5. **CPS** の方向
6. まとめ

「シミュレーション研究」の過去・現在・未来

大雑把なイメージとして、…

- **60年前(1960):** プログラムサイズ **4000** 行くらい(カードを一人でもてる上限)、メモリサイズ **64k** ワード、計算速度 **1Mflops**
- **30年前(1990):** プログラムサイズ 最大**10** 万行とか?、メモリサイズ **32GB**、計算速度 **100GF**
- **今(2020):** プログラムサイズ **??**、メモリサイズ **5PB**、計算速度 **500PF**

(このクラスの計算機の数全然違うので、演算能力はそこまで増えてない)

- 良いこと: **11** 桁以上の計算能力の向上。いろんなことが(少しは)計算できるようになってきた感。
- 良いともいいがたいこと: プログラム開発、保守の困難さの増大

プログラム開発、保守の困難さの増大の理由

- モデルが複雑になった
 - 複雑な物理をいれられるようになった(輻射、化学反応等)
- スキームが複雑になった
 - 高速化・高精度化のための複雑な計算法
 - 解適合格子等のアダプティブ法
- 計算機が複雑になった
 - 分散並列、ノード内並列、階層キャッシュ、**SIMD** 演算ユニット、アクセラレータ、、、

どれをとっても、1つだけでも1人でできるかどうか、という程度に複雑

宇宙物理における大規模シミュレーション ソフトウェアの現状

- 大規模な **MPI** 並列コードでも (超人的な) 個人が作ってメンテナンスもしているのが多い
 - **Gadget**、**ASURA** (銀河形成 **N**体+**SPH**)、**Sfumato AMR** (星形成 **AMR** 流体)、**Rebound**(リング等 **N**体) (**NBODY6++GPU**) ...
 - 物理モデルはプラグイン的(でないものもある)
- **MPI** 並列化できてないものは最近でも論文レベルで無数にあるがプロダクションランに使えるのか?
- 主に大学ベースのグループでの開発もある: **Enzo**, **RAMSES**, **Athena++**, **pkdGrav/Gasoline**, ...
- 他の目的ともつながりのある大規模な開発も: **FLASH**
- 物理モデルはくっつける感じに。(化学反応ネットワーク、輻射、恒星進化等)

現状の整理と問題点

- 分業がないわけではないが、「力学」と「物理」的な分業。言い換えると「数値計算スキーム+計算機対応」と「物理過程」。
- 「計算機対応」が非常に困難になっている
- この**30**年で計算機がものすごく使いにくくなった
 - 共有メモリベクトル並列
→分散メモリマルチコア階層キャッシュスーパーカラー**SIMD**
 - 演算カーネルの部分の最適化だけでもマシンによっては極めて困難で、しかも実現できる効率が低い(例:**A64fx**)
 - 階層キャッシュの構造が数値スキームに適していないため原理的に性能がでないケースが多い。
 - (計算機の開発方向が間違っている)

我々の対応

理研 **AICS/CCS** 「粒子系シミュレータ開発チーム(今年度でめでたく終了)でやってきたこと:

「並列化」「計算機アーキテクチャ向け最適化」「数値スキーム」の分離

- 並列化: **FDPS** が担当。計算領域のロードバランスを考慮した分割、粒子の領域間の交換、領域にまたがった相互作用計算のための通信(重力・クーロン力のような長距離力も)
- アーキテクチャ向け最適化: **PIKG** が担当粒子系で最も重要なのは相互作用カーネル。高レベル記述から **SIMD** 命令を使ったループや **Cuda** カーネルを生成
- 数値スキーム: 時間積分とかはユーザーが書く

FDPSの成果

- そこそこ色々な人に使われている。**2016**年の論文で今までの引用**82**なのでまあまあ。
- **FDPS**を使った大規模並列コードもいくつか。**PETAR, PENTACLE, GPLUM, ...**
- 今まで大規模並列コードがあんまりなかった業界で大規模並列計算を可能にする、という目標は結構達成できた

FDPSの今後

- 富岳で性能低いところ改善(この**3**ヶ月でだいぶましになった)
- 長距離相互作用の計算アルゴリズムの改良 (**FMM**, **PM³** の実装)
- いわゆる **FMM** でのエネルギードリフト問題への対応

粒子法以外は？

格子法:

- **AMR** とかでない単純な規則格子なら原理的には差分スキーム自体の表現からの **MPI** 並列コードの生成はそれほど難しいわけではない
- キャッシュ使って性能だそうと思うと複雑なアルゴリズムが必要だが、現行の階層キャッシュではそれで性能あげるのは非常に難しい。今までいくつかのマシンで実験して効果あったのは **PEZY-SC2** だけ。 **Sunway Taihulight** もちゃんとやれば効果あったかも。京、富岳は上手くいってない
- 規則格子で、普通に人が書くのに比べて速くないなら、、、でちょっと止まってる感じ。

(**AMR**: 個人的にはメッシュレス粒子法のほうが幸せになれるのではという気も)

「キャリアパス」問題

コード開発やっててサイエンスの論文がないとポストがない問題

- 古典的な解決: 超人になってサイエンスの論文も書く
- 古典的な解決の問題: 多くの人は超人ではない
- 素粒子的な解決: すべての人が論文に名前を並べる
- 素粒子的な解決が機能する(というかこの解決にならざるを得ない)条件: 目標が**1**つしかない。宇宙でこれに近い: 重力波、**EHT**(ブラックホール降着円盤解像)
- 宇宙科学・惑星科学で素粒子的な解決が機能しない理由: 研究対象が**1**プロジェクトでも沢山あってすべての論文にすべての関係者が名前並べるわけにもいかないから?

論文がありさえすればいいのか？

- 正直なところそうかもしれない。
- コード開発は知的生産であり、そこでの新しい知見は論文になるべきである。
- 「こんなことは論文にならない」というのは「我々の」偏見にすぎないかも。

40年前だと、「数値計算アルゴリズム」そのものは天体力学では論文になってもそれ以外の宇宙物理では必ずしも論文にならなかった。少なくとも **Astrophysical Journal** 本誌にはのらなかつたりした。有名な例: **Barnes and Hut 1986, Nature, 324, 446-449 A hierarchical $O(N \log N)$ force-calculation algorithm**

- **D** 論は装置じゃなくて観測で、とかシミュレーションアルゴリズムだけでは **D** 論じゃないとかはまだそこらじゅうにある気もする。
- (とはいえ、命令セットやキャッシュ構成が××な××をがんばってこれくらい使えるようにしました、は本当に知的生産なのかという気も)

惑星科学における探査研究との関係

- 私あんまりわかってない。
- 天文でも、衛星ミッション計画やデータ解釈にシミュレーション必要というのはある。
- 計画段階のは大変な割にこっちの研究成果にならなかつたりする。
- 最近のようにシミュレーションソフトウェア開発が大変だと一層これはしんどそう。
- この辺も「論文化」の問題かも

CPS の方向

ここまで一般論。ここは **CPS** スペシフィックな話。

これまで(というかここ **6** 年くらい)の **CPS**

- 惑星探査コンソーシアムのハブ機能 (宇宙研連携)
- シミュレーションソフトウェア開発基盤
- 京・富岳での計算惑星科学

今後: 計算科学的な方向に (組織維持のためにも研究費も必要、、、)

- シミュレーションソフトウェア開発基盤
- 京・富岳での計算惑星科学
- 計算機開発、データ (惑星) 科学、**AI**、、、
- 惑星探査コンソーシアムにもこの方向から貢献を

まとめ

- 宇宙科学・惑星科学のシミュレーションソフトウェアの開発は困難な時期を迎えている
- その大きな要因は、計算機ハードウェアの複雑化によりまともに性能がでるアプリケーションを開発するのが非常に大変な作業になったためである
- このため、なかなか論文にならない作業が膨大になった
- 観測・探査ミッションへの貢献においても、このことに起因する問題もある
- この辺、「科学的成果」とはなにかの考え方を変える必要があるかもしれない