

次世代スパコンへの期待

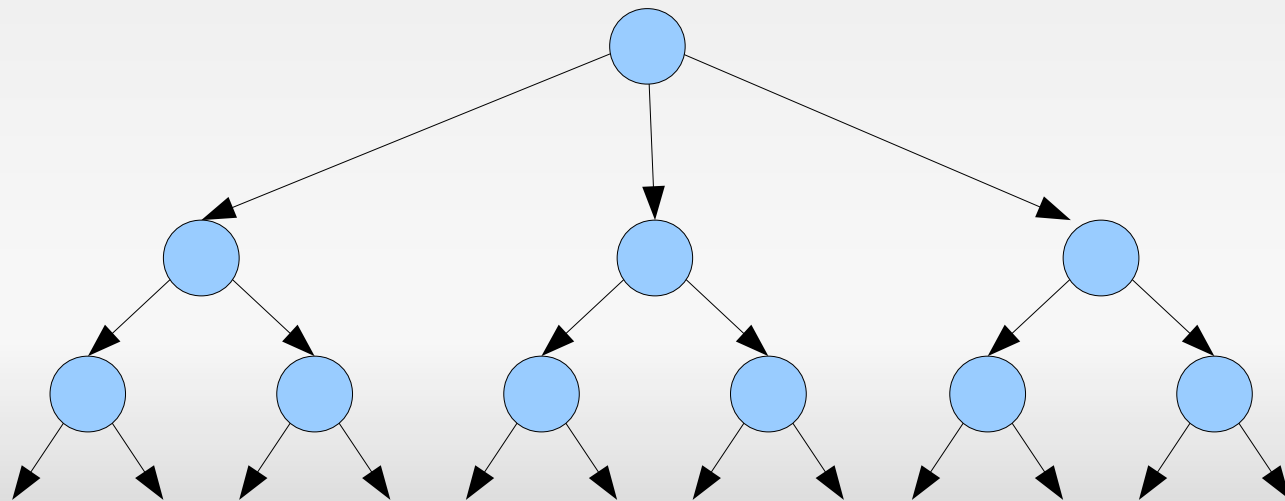
- 田浦健次郎
- 東京大学 情報理工学系研究科
情報基盤センター（兼務）

個人的バックグラウンド

- 1993- 並列 OO 言語 ABCCL on 富士通 AP1000+
- 1997- 超軽量マルチスレッディング (StackThreads/MP), 並列 C/C++ 言語
- 1996- 分散 / 並列 GC
- 2005- InTrigger cluster of clusters
- 2004- 手軽な並列分散処理ツール GXP, GXP make
- 並列処理を高水準に (簡単に , 楽しく) するのがテーマ

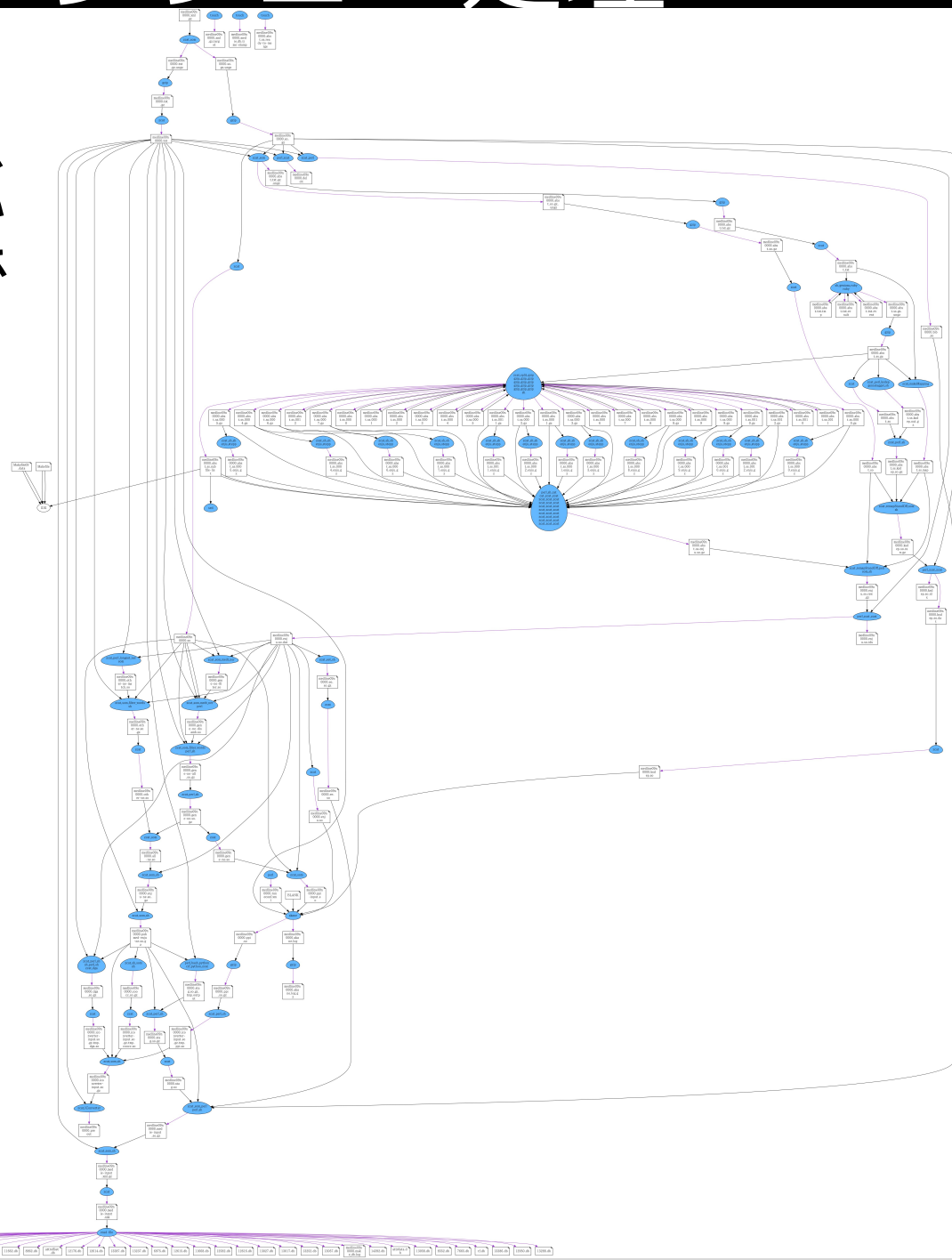
超軽量マルチスレッディング

- なんのためのもの？ 「階層的，動的に生まれる」並列性を自然に記述できるもの
 - cf. OpenMP の tasks, Java forkjoin, Cilk, etc.
 - 分割統治・探索問題などに有効 (MPI で管理するのは困難)
- 簡便な記述 + 自動負荷分散 + ローカリティの高い実行



GXP の手軽なワークフロー処理

- 既存コンポーネント (バイナリ) の組み合わせで行う並列処理
 - ゲノム解析, 画像解析, 自然言語処理, etc.



誰でも並列プログラミングの時代？

- クロックの向上，単体 CPU(逐次) 性能向上の限界 ⇒ マルチコア
- 性能向上のためには，ソフトウェアによる並列処理 (= 並列プログラミング) がこれまでになく広い場面で必要になっていく
 - 「デスクトップ並列処理」を今や誰もが当たり前と思っている
 - それは MPI (だけ) のことではない
 - 並列処理の裾野は広がっている
 - MPI は知らなくても CUDA, EC2, MapReduce は知っている

並列処理・プログラミング小史 (1)

■ 夢多き時代 (~ '90 前半):

- 様々なアーキテクチャ (MPP, データフロー, etc.), 言語 (自動並列, 関数型, 論理型, オブジェクト指向) が百花繚乱
- さもなくば並列プログラムなど作れないという考えで理想を追求

■ コモディティ・標準化時代 (90 後半 ~):

- MPI, PC クラスタ, Linux ベースなど, 並列処理の標準フォーム ⇒ HPC 普及・産業利用の下地
- 一方「新言語」には冷たい視線

並列処理・プログラミング小史 (2)

■ 現在

- アーキテクチャ・OS は再び多様化 (試行錯誤中?), 非 commodity 化している
- ヘテロ, 縮退したノード OS
- 一方, 並列処理の分野・裾野は広がっている
 - 両者の断絶は心配

今後の並列処理 (1)

- **高生産性**への要求の高まり
 - cf. High Productivity Languages (X10, Chapel)
 - 並列処理人口, アーキテクチャの複雑さが増すにつれ一層重要になる
- **「大きな」**並列処理
 - 複数の並列タスクや, 多数の逐次コンポーネント, を多数組み合わせた大きな並列処理
- **「実データ処理」**中心の並列処理
 - ゲノム, 自然言語処理, ウェブ解析, etc.

今後の並列処理 (2)

- 移植性への要求の高まり
 - 異世代マシン間の移植性
 - マシンの寿命 < アプリの寿命
 - 同世代マシン間の移植性
 - デスクトップ～中規模マシン～大規模マシンまでの連続性
 - 産業利用を支える，または産業利用に支えてもらうためにも重要
- ソフト = アプリのことだけではない．ミドルウェア，言語，ツール etc. の移植性を無視してはいけない！
 - 基盤ソフトウェア屋は「そこでしか動かない」ソフトに対する興味がない

計算機科学と計算科学のよい協調関係を築くために (1)

- 計算機科学にとっての win
 - 新しい方式 (言語 , ツール , ライブラリ) が息長く使われていくこと
 - ≠ 「このマシン用のコンパイラ」
 - 将来性のないマシン (アーキテクチャ) 用の方式はなかなか作れない
 - 陥りやすいピットフォール : 「このアプリと新マシンの間をソフト (コンパイラ , ライブラリ) が埋めてくれる」という期待

計算機科学と計算科学のよい協調関係を築くために (2)

- 今は「並列ソフトの書き方を変えなくては」という機運が高い，チャンスな時代と言える
 - GPU 付き，マルチコア，何万ノード並列ソフトをどう書けばいいのか誰も分からない
- しかし計算科学者から見ても将来性のない，普及していない言語はなかなか使えない
- 計算機科学者が計算科学を勉強し，アプリを盗むことが重要
- 計算科学の人々がそれ（アプリの書き直し）を微笑ましく見守る事も重要

計算機科学と計算科学のよい協調関係を築くために (3)

- 探索問題など動的・階層的並列性
- 極めて高並列な (MPI で苦勞する必要ない!), モンテカルロ的アルゴリズムの超簡便な記述
- 環境を発展させる役割・貢献
 - 「次世代コンピュータで提供されるソフトウェアはこれです (Fortran, C/C++, MPI, OpenMP, ...) 」というのはおかしい

次世代スパコンへの期待

- 我が国の最高性能マシンとして君臨するのみならず，
- 同モデルが研究用，産業用に「買える」マシンとなること（を目指すこと）
- そして，ソフトウェア研究・開発者にとって，魅力的な研究・開発対象となること
- それにより，次世代スパコンという「プラットフォーム」が息の長いソフトウェア研究・技術の集積場になることを期待

技術蓄積のためのモデル作り (1)

- ソフトウェアは HPC で利用されているものを含めて多数がオープンソースや無料で提供されている
- 技術を開発すればそれを商品化して「売れる」わけではないところにソフトウェアの特殊 (?) 事情がある
- 「エコシステム」構築の困難さ

情報技術蓄積のためのモデル作り (2)

- この状態で新規ソフトウェア技術開発へ人をどう惹きつけるかは極めて深刻な課題（情報系学生はGoogle 志望者ばかり）
- 米国：
 - 大学での研究 → ベンチャで開発 → 大企業が買収 というサイクル
 - 大企業のオープンソースソフト・プログラマへの投資
 - 例：Lustre ファイルシステム，仮想マシン，...
 - 大学が研究成果を実用的レベルのソフトウェアとして提供することへの意欲も高まる

計算科学研究機構への期待

- 次世代，次次世代，その先，... 情報技術の「蓄積」（基盤技術を失わないこと）を目的としたコラボレーションの場
 - 計算科学研究者と組んで MPI 大規模並列化を支援するのはもちろん重要
 - しかし，成功の定義はそこにとどまるべきではない
- 並列計算の「裾野」が広がっている事を意識した（ハードコアな HPC に止まらない）研究コミュニティ作り

情報基盤センターの役割（私見）

- 次世代スパコンと連携することで，大学が果たすべき役割を今よりも果たしやすくなる
 - 一言で言えば「おおらかなマシンの使い方ができるようになる」
 - 人材養成
 - その先の基盤ソフトウェア研究

まとめ

- 次世代スパコンが一世代に止まらず，息の長い情報技術蓄積・計算機科学と計算科学の協調のプラットフォームとなることを期待
- 並列計算の裾野が広がっている事をむしろ生かしたい
 - 難しい高性能プログラミングが出来る人以外お呼びでない，という感じではいけない
 - フラッグシップだけでなく中規模，小規模マシンを買えるようでありたい