

SWoPP鳥取2012 参加報告

大島聡史, 實本英之, 鴨志田良和, 片桐孝洋, 田浦健次郎, 中島研吾

東京大学情報基盤センター

1 SWoPP鳥取2012

本記事ではSWoPP鳥取2012(以下、SWoPP2012)について報告する。SWoPPは「並列/分散/協調処理に関するサマー・ワークショップ」(Summer United Workshops on Parallel, Distributed and Cooperative Processing)の正式名称を持ち、1988年のミニシンポジウムに端を発し今年で25回目を数えるワークショップである。主催団体は電子情報通信学会・情報処理学会・日本応用数学会の3学会8研究会/研究部会である。今年は鳥取市にて8月1日から3日の日程で開催された。

SWoPP2012では参加者は300名を超え、合計138件の発表(研究会/研究部会による招待講演を含む)が最大4並列で行われ、さらに3件のBoFが開催された。世間がロンドンオリンピックに盛り上がる中、鳥取では体温を超えるような夏の暑さにも負けず、連日熱い議論が繰り広げられた。

情報基盤センターの教員から(教員が第一著者であるもの)は以下の6件の発表が行われた。(タイトル末尾角括弧は発表された研究会/研究部会の略称。)

1. 中島研吾(東大): Oakleaf-FX (Fujitsu PRIMEHPC FX10)における前処理付反復法の最適化 [MEPA]
2. 伊藤祥司, 杉原正顕(東大): 前処理付きCGS法の改善とクリロフ部分空間に注目した解析 [MEPA]
3. 田浦健次郎, 中島潤(東大), 横田理央(KAUST), 丸山直也(理研計算科学研究機構): ExaFMMのタスク並列処理系MassiveThreadsによる並列化とその評価 [HPC]
4. 鴨志田良和(東大): クロスコンパイル環境でインストール時チューニングを容易にするミドルウェアの開発 [HPC]
5. 實本英之(東大): 通信オーバーラップを考慮したアプリケーションレベルチェックポイントフレームワーク [DC]
6. 大島 聡史, 實本 英之, 鴨志田 良和, 片桐 孝洋, 田浦 健次郎, 中島 研吾(東大): 大規模超並列スーパーコンピューターシステムOakleaf-FX(FUJITSU PRIMEHPC FX10)の性能評価 [HPC]

本記事ではSWoPP2012に参加した複数教員の視点から、各々が興味を持ったセッションや研究について概要や動向の紹介を行う。

2 GPU

画像処理用のハードウェアであるGPUを用いて高性能な計算を行うGPGPUは近年のHPCの分野において注目を集めている研究テーマである。SWoPP2012では昨年のSWoPP2011に引き続き複数の研究会/研究部会セッションにおいてGPUに関する発表が行われた。

GPUに関する研究は様々な分野に広がりを見せており、今回のSWoPP2012においては、数値計算、動画像処理、気象、暗号化、文字列探索、CPU-GPU間の通信最適化、プログラミング環境と様々な研究報告が行われた。全ての研究がGPUに対して最大限に最適化を行い劇的な性能向上を得られているわけではないが、今後もさらに広い分野にて研究が行われ、利用されていくだろう。

GPUに関する多くの研究はNVIDIA社のGPUとCUDAを用いて行われていたが、XMP(XscalableMP)のGPU対応やOpenACCを用いたものなど、CUDA以外の(アプリケーションプログラマがCUDAを直接記述しない)GPUプログラミングに関する研究についても報告があった。特に現在では指示文を用いて容易にGPUプログラミングを行えるOpenACCへの期待が高まっており、今後の研究会等にてOpenACCに関する研究発表が多く行われると考えられる。

3 DC

各種コンピューティングシステムにおける耐故障性をあつかうDC(ディペンダブルコンピューティング)研究会では、3件の発表と1件の招待講演が行われた。

1件目の発表は、ディレクティブベースのチェックポイントフレームワークに関するもので、アプリケーションレベルチェックポイントにおいてアプリケーションプログラマが考慮することが難しい故障率といった環境情報をライブラリ・コンパイラのサポートにより加味することで、最適なチェックポイント周期を提供するというものであった。

2件目の発表は、P2Pネットワークにおいて一部ノードのリンク情報が何らかの要因で不完全となった際にどのような影響があるかを解析したもので、目的ノードまでの経路情報の取得を目的とした際のノード数の増加による影響や代替リンクの量による影響を考察したものであった。

3件目の発表は、耐故障機能が不全になる事によって起こる二次リスクをどのように保証すれば良いかという検討と、実際の事例からの経験を報告したものであった。最終的には全ての事例をデータフローダイアグラムで示し、末端に証跡(仕様書の記述やテスト実行の結果等)を記述しておくことで保証ケースを万人に納得できる形で示すことが可能という内容であった。

最後の招待講演は、近年のクラウドコンピューティングの運用・導入コストや運用法、省電力化の実例や故障事例からクラウドにもディペンダビリティが必要であることの示唆、及び、どのようなディペンダビリティが必要かの検討を行ったものであった。最終的には、上記で述べた事例のパラメータに鑑みながらディペンダビリティの高いシステムを構成するためのますますの研究を励行するという内容となっていた。

DC研究会では大規模高並列環境におけるユーザプロセスの耐故障性の話から、より大きな運用不具合発生時のサービス停止の局所化、ディペンダビリティと省電力化の兼ね合いといった広範囲に適用できるような発表が行われており、センターのサービスを考える上でも有用なものであった。

4 自動チューニング

自動チューニングは、実行環境のCPUやメモリの性能、実行時の並列数、入力データなどの情報を活用して最適な実行パラメータを探索し、コンパイラの静的な最適化よりも高い性能を得

るための技術である。HPC研究会ではこの自動チューニングに関するセッションが設けられ、3件の発表が行われた。

1件目は行列の固有値ソルバの最適化に関するもので、ノード内のCPUコアを通信制御専用にする事で非同期集団通信を実現し、これによって得られる通信コストの隠蔽の効果と、計算に利用されるコアが減少することによる計算時間の増加のトレードオフを考慮した最適な通信オーバーラップの方法を、自動チューニングによって決定するという内容であった。

2件目は自動チューニング機能付きソフトウェアの導入支援をするツールに関するものであった。京やFX10においてクロスコンパイラを実行したホストから生成されたファイルを実行しようとした時に、それを自動的に計算ノード上のジョブに変換してから実行することで、自動ソースコード生成、コンパイル、実行という、自動チューニング機能付きソフトウェアで頻出する一連の処理を透過的に実施できるようにするという内容であった。

3件目は正方向行列積計算の性能について、ある行列サイズ m での性能値から、別の行列サイズ n での性能を推定するというものであった。自動チューニングでは最適な実行パラメータを見つけるために、多くのパラメータを探索する必要があるが、精度の高い推定が可能になれば探索コストを減らせるため、重要度が高い。

3件の発表のうち、2つは京やFX10に関連するもので、その注目度の高さが伺える。自動チューニング技術がもたらす実行環境に適応した最適化やその自動化は、計算の大規模化に伴って益々重要になっていくものと思われる。

5 MEPA

MEPAは日本応用数学会「行列・固有値問題の解法とその応用」研究部会が主催する研究会である。SWoPPには、著者の一人である片桐が同研究会の幹事である時期(SWoPP佐賀2008)から参加している。今年度は3セッション9名の発表があり、そのうち1件は招待講演であった。

「疎行列反復解法の前処理」のセッションでは、東大のFX10(Oakleaf-FX)やHA8000(T2K東大版)を用いた数値計算アルゴリズムや自動チューニング機能付き数値計算ライブラリの発表が行われた。

「アプリケーション適用」のセッションでは、大規模電子状態計算への応用や、価格時系列相関の固有値分布を用いた株式市場の主要産業セクタ抽出といった、応用分野の発表が行われた。

最後の「高精度演算」のセッションでは、高精度演算や多倍長演算関連の発表が行われた。このセッションでの最後の発表は招待講演であり、静岡理工科大の幸谷智紀先生による「多倍長疎行列・ベクトル積の実装とクリロフ部分空間法への応用」の講演が行われた。スーパーコンピュータを用いた数値計算から数値アルゴリズムまでの広範な分野を含む発表が行われた。

MEPAセッション参加者の中には日本応用数学会に所属していない参加者も見られ、SWoPPの目的である他分野との連携がなされていると言える。

6 京コンピュータ・FX10に関連した発表

京コンピュータ、FX10に関連した発表は上述のセンター教員によるFX10に関する2件の他、京コンピュータ上での並列プログラミング言語XcalableMPの評価(村井(理研計算科学研究機構)他)、性能解析ツールScalascaを用いた性能評価(中村(理研計算科学研究機構)他)についての報告があった。今田(富士通)らによる「ハードウェアプリフェッチ機構の活用によ

るステンシル計算の性能改善手法」では、気候系アプリケーションに見られるようなステンシル計算のカーネルループを対象として、ハードウェアプリフェッチ機構を有効活用することによる性能改善手法が提案されている。並列計算における袖領域の存在によって生じる不連続なメモリアクセスを連続に見せるようなオペレーションを適用して、ハードウェアによるプリフェッチ動作の継続を促すことで、SPARC64 VIIIfxにおいて12%の性能向上が得られたという報告があった。

7 高水準並列処理

現在多く用いられているMPI, MPI+OpenMP, MPI+CUDAのような並列プログラミングモデルよりも、並列化を簡単に、かつ可搬性の高いものにする試みがなされている。軸となるアイデアは主に2つで、一つは、各ノード毎に分離されたアドレス空間ではなく、全ノードにまたがったアドレス空間を提供するプログラミングモデルPGAS (Partitioned Global Address Space)、もう一つは、プログラム実行中の任意の場所でタスクを生成し、それを処理系が負荷分散するタスク並列処理である。HPC研究会においても関連する発表が多くあった。

筑波大学 佐藤らを中心として開発されているXcalableMPに関して、その京コンピュータ上での評価、高性能なLinpackの実装および評価について発表があった。128プロセス時にMPIで書かれたHPLと比べて70%程度の性能を持ち、コードの行数は半分程度ですんだと報告されている。

タスク並列については、筆者の一人である田浦らによる、軽量スレッド処理系(MassiveThreads)およびそれを用いたアプリケーションスタディに関する発表が行われた。処理系については、負荷分散方式をアプリケーションごとにカスタマイズするためのAPIおよび、分散メモリ環境でのタスク並列処理モデルのためのPGASのAPI設計に関する発表が行われた。アプリケーションスタディとして、タスク並列処理系を用いたFast Multipole Method (FMM)およびAdaptive Mesh Refinementの並列化およびその評価に関する発表が行われた。どちらも、計算量やメモリ量削減のための木構造の生成やそれを用いた計算を再帰呼び出しを用いて記述したものでタスク並列によってそれを見通しよく並列化できることが示されている。

また、スクリプト言語を用いて比較的粗粒度なタスクを多数並列処理する、並列スクリプティングやワークフロー処理系に関する発表が3件あった。筑波大学 辻らによる、XcalableMPプログラムとタスクとしてそれを多数並列実行するためのワークフロー言語YMPに関する発表があった。また筑波大学 田中らの発表は、RubyのRakeを元にした並列ワークフローPwrakeに関するもので、100万ノード/1億タスク規模のワークフローを実行する際の、主にメモリサイズから来る限界を指摘して、それを解消する方法についてであった。また、三重大学 仲らによる、MegaScriptにおける通信の効率化に関する発表があった。

スケラビリティに関して言えば、タスク並列にせよ、スクリプティング・ワークフローにせよ、「階層性」をうまく活用することが鍵であるという方向性は多くの研究に共通している。タスク並列処理では、アルゴリズムを再帰的に記述することで、処理系は負荷分散を柔軟かつ局所性をこわさずに行うことができる。ワークフローについてはタスクをディスパッチする部分のボトルネックを解消するために、ワークフロー記述にタスクのグループ化を明示的に取り入れる、などのアイデアがそれにあたる。

スクリプトやワークフローなどの粗粒度な並列処理は、大規模計算機の「キャパシティ」を

活用するための有用な方法で、かつ実用的にも重要である。個々のプログラムのスケーラビリティを目指すことはチャレンジングかつ重要であるが、実際の応用ではプログラムを一度だけ走らせて結果が出るということはない。実際には並列プログラムの「外側」に、そのプログラムを様々な初期条件やパラメータで走らせるという疎粒度な並列性がある。このような疎粒度並列処理をうまく制御して、最終結果を短時間で(例えば最適なパラメータや初期条件を探索することで)得ることができれば、これも一種の階層性の活用である。

次回のSWoPPは2013年夏に小倉にて開催される予定である。