# 研究課題別事後評価結果

1. 研究課題名： 超低電力化技術によるディペンダブルメガスケールコンピューティング

2. 研究代表者名及び主たる研究参加者名（研究機関名・職名は研究参加期間終了時点）

研究代表者

中島　浩 （京都大学学術情報メディアセンター　教授）

主たる共同研究者

中村　宏 （東京大学先端科学技術研究センター　助教授）

佐藤　三久 （筑波大学計算科学研究センター　教授）

朴　泰祐 （筑波大学計算科学研究センター　教授）

松岡　聡 （東京工業大学学術国際情報センター　教授）

3. 研究内容及び成果：

本研究の目的は

(1)　ハードウェア／ソフトウェア協調による低電力化技術

(2)　大規模並列タスクの実行モデル構築・利用技術

を柱として、種々のコモディティ技術を活用したメガスケールコンピューティングの基盤技術を確立することにあった。すなわち、この 2 つの技術を中核としてプロセッサ、コンパイラ、ネットワーク、クラスタ管理、およびプログラミングに関する研究を行い、それらにより 100 万プロセッサ級の汎用メガスケールコンピューティングが実現できることを示すことと、そのプロトタイプとして

(3)　低電力・高密度大規模クラスタ MegaProto

を構築して技術の有効性を実証することが、本研究の目的であった。

(1)　ハードウェア／ソフトウェア協調による低電力化技術

この技術の確立を目指す研究は、以下の2グループにより実施した。

【研究項目1】　ソフトウェアとの協調最適化に基づく超低消費電力技術・高密度実装技術・高バンド幅技術：プロセッサグループ（リーダ：東大・中村）

【研究項目2】　ハードウェアとの協調最適化に基づき低消費電力かつ高性能を実現するコンパイラ技術：コンパイラグループ（リーダ：筑波大・佐藤）

研究の鍵となるハードウェア技術は、SCIMA (Software Controlled Integrated Memory Architecture)と呼ぶ、ソフトウェアから可視かつ構成の変更が可能な高速メモリ階層アーキテクチャである。SCIMA は、通常のキャッシュとの境界が可変である高速メモリ SCM を中心に構成され、配列などのデータは再利用性、アクセスの規則性、容量に応じて SCM あるいは通常のキャッシュ可能な空間に割付けられる。この割付けをコンパイラが最適化することにより、プロセッサチップと主記憶の間のデータ転送の回数や量を大幅に削減することができ、さらにオンチップメモリのアクセスによる消費電力も削減できる。この結果、実行時間と消費エネルギーの両面で、大きな削減効果が達成された。

一方、従来型のプロセッサについても、ハードウェア機構、特に電源電圧や周波数を動的に変更する DVS 機構を、最適化コンパイル技術により活用して、消費電力を削減することができる。本研究ではこの DVS とコンパイラの協調技術を多角的に追求し、プログラムフェーズごとの電力プロファイルに基づく最適化、負荷不均衡プロセス群の電圧・周波数最適設定、実効消費電力の動的制御による性能最適化、性能カウンタの統計情報に基

づく電力最適化といった技術を提案し、いずれも優れた効果を得た。

(2) 大規模並列タスクの実行モデル構築・利用技術

この技術を目指す研究は、以下の3グループにより実施した。

【研究項目3】 安価かつスケーラブルな高速・高信頼ネットワーク技術：ネットワークグループ（リーダ：筑波大・朴）

【研究項目4】 実行モデル構築とモデルを利用した大規模コモディティクラスタ管理技術：クラスタ管理技術グループ（リーダ：東工大・松岡）

【研究項目5】 メガスケールのタスク並列プログラミングとその実行技術：プログラミング技術グループ（リーダ：豊橋技科大／京大・中島）

実行モデルの構築のために、並列タスクの挙動情報を記述可能なタスク並列スクリプト言語 MegaScript を設計した。この言語ではコンパイラの解析情報からは決定困難なタスク挙動に関する量的情報を与えることができ、静的あるいは動的な実行モデルを高精度に構築することができる。生成されたモデルは、MegaScript で記述された並列タスク実行のスケジューリングのために用いられ、タスク粒度の調整や最適な配置が行われる。

また、実行モデルの別の重要な応用として、システムの信頼性向上に関する研究も行った。この研究では、大規模システムの各ノードで実行される、挙動が類似したプロセスの関数呼出／復帰のトレースにより挙動モデルを構築し、モデルから大きく外れたプロセス挙動を異常なものとして統計的に分別する。またこの手法を 129 ノードの大規模クラスタに適用して、長期間の運用中に生じたシステムダウンの原因を解析した結果、システムのミドルウェア中に間歇的に生じる異常の原因となる2つのバグを発見することができた。

また大規模システムにおける脆弱性の主要因であるネットワークの耐故障性のために、ノード間リンクやスイッチ間のパスを多重化して高バンド幅と高信頼性を同時に実現する、RI2N (Redundant Interconnection with Inexpensive Network)と、VFREC-Net (VLAN- based Flexible, Redundant and Expandable Commodity Network)を提案し、MegaProto 含む種々のクラスタに実装して性能・信頼性の両面での有効性を実証した。さらにハードウェアレベルの耐故障機能や、チェックポイントの生成・回復機能を統合した、耐故障 MPI フレームワークである Cuckoo を開発した。

(3) 低電力・高密度大規模クラスタ MegaProto

多数の低電力プロセッサを高密度に実装し、それらを高信頼・高バンド幅のネットワークで結合したプロトタイプシステム MegaProto を開発した。また MegaProto は、プロジェクトで研究・開発中の様々な技術の実証プラットフォームとしても利用された。

MegaProto の仕様設計は5名のグループリーダを中心とした設計チームを組織して行い、TM5800 (Crusoe)を用いたテスト機（MegaProto/C）と、TM8820 (Efficeon)を用いた実用機(MegaProto/E)の詳細な仕様を定めた。この仕様に基づき実装設計および製造を外注し、MegaProto/C を2台(32PE)、また MegaProto/E は20台(320PE)をそれぞれ製造した。これらの MegaProto システムはいずれも優れた電力性能比を達成し、特に MegaProto/E の 100MFLOPS/W はコモディティ技術を用いた並列システムとしては世界最高の値を達成した。

4. 事後評価結果

4－1. 外部発表(論文、口頭発表等)、特許、研究を通じての新たな知見の取得等の研究成果の状況

前半は論文発表数が少なかったが、研究成果の出始めた研究後半になって発表数が増加した。大規模なシステムを実際に作成するという研究の性質上これはやむを得ない。情報処理学会誌などの国内論文 15 件、IEICE の国際論文 1 件である。口頭発表は Cool Chips や International Symposium on High Performance

Computing 等の国際会議 21 件、国内会議 17 件、また、スーパーコンピューティング分野の代表的且つ世界最大の国際会議である Super Computing2005、2006 に投稿して採録されるとともに、作成したシステムの展示を行うことにより、世界での認知度が高くなっている。国内でも、この分野での代表的なシンポジウム SACSIS や情報処理学会論文誌で多くの論文が発表され、この分野の主要な地位を占めている。しかしながら、成果として存在するものの未だ国際会議で戦っていない部分もあり、国際的にみれば論文数はもう少し多くあってしかるべきであろう。

具体的な特許は出されていない。この分野では、システムの実際的な製作によって始めて知見が積み重なるという性質上、この分野の特許取得は困難である。しかしながら、技術の切り出し方によっては特許にすることのできる要素も幾つか有ったように思われる。他方、多くの有用なソフトウエアが開発され、著作権のある形で存在している。

研究内容としては、要素となる諸技術を定義し、着実な仕事によりそれらを実現した。SCIMA アーキテクチャ、電力プロファイル駆動方式、相互接続ネットワークでは RI2N 方式の実装と故障縮退方式、クラスタ管理手法ではモデルベースのソフトフェイリャー検出方式、プログラミングでは処理系のスクリプトによるユーザレベル拡張方式など、それぞれ有意義な研究成果が出ている。また、これらを組み込んだプロトタイプシステムとして MegaProto を 320 台のプロセッサで実現し、しっかり動作させ、その性能が高いことと、コモディティクラスタとして世界最高の性能電力比を実現した。

4－2. 成果の戦略目標・科学技術への貢献

汎用部品を用いて高性能なシステムを低電力で実現するという手法を提案してその実現方式を与えた。これには様々な要素技術が関係しており、それらの要素技術の新規性とともに、科学的・技術的なインパクトは高い。特に、ネットワーク技術、キャッシュとローカルメモリの動的なバウンダリー化技術、高性能計算における低電力技術、耐故障技術、モデリング技術などがそれで、この分野の研究は世界でも精力的におこなわれているが、その先駆けとして優れた成果を出した。また、当初は余り考えられていなかった性能予測手法のチューニングが進み高精度な並列計算モデルを実現した。これらの成果は、戦略目標としての、今後求められる超高性能なスーパーコンピュータの重要な要素技術を与えるものである。

ここで開発した諸要素技術は、システム技術として今後様々に利用され発展することが考えられる。現に、幾つかの大学の新プロジェクトに於いてそれらを核として発展させる研究が始まっている。また、成果のソフトウエア自体、多くは実用レベルに達しており、それによって基盤環境が揃い、その活用による新たな実験と展開が期待される。また、今後省電力化プロセッサとして、マルチコアプロセッサの出現が見込まれるが、本研究でなされたマルチコアネットワーク技術やオンチップメモリ技術などは、それにも大いに適用できるものである。

4－3. その他の特記事項（受賞歴など）

要素技術の開発と、それらを組み合わせたプロトタイプの実現がこの研究作業である。その総合的なプロトタイプが成功裡に早期に実現できたことは、研究体制が適切であったことを示している。地域的に分散した 4 大学にまたがるプロジェクトであったが、研究代表者の強い求心力で円滑に研究が推進された。当初計画の中心課題の維持と、諸テーマの取捨選択、新規テーマの発掘など順調であった。

世界最高レベルの国際会議 Super Computing において、3 年連続して研究展示と発表を行うとともに、国内の学会で最優秀論文賞など 3 件の受賞を受けた。また、高性能計算分野で代表的な世界の研究機関、Argonne National Laboratory、University of Wisconsin、INRIA などとの間で活発な研究者間の交流が行われた。