次世代超並列計算機開発
— 連続体向け超並列計算機の開発

Next-Generation Massively Parallel Computers
— Development of Massively Parallel Computers for Continuous Physical Systems

岩崎洋一 (筑波大学計算物理学研究センター)

1　研究組織

プロジェクトリーダー	岩崎洋一	筑波大学物理学系教授 素粒子物理学、全体とりまとめ
コアメンバ	宇川彰	筑波大学物理学系教授 素粒子物理学
	金谷和至	筑波大学物理学系助教授 素粒子物理学
	青木慎也	筑波大学物理学系助教授 素粒子物理学
	吉江友照	筑波大学物理学系助教授 素粒子物理学
	梅村雅之	筑波大学物理学系助教授 宇宙物理学
	中本泰史	筑波大学物理学系助手 宇宙物理学
	大川正典	高エネルギー加速器研究機構数値物理部助教授 素粒子物理学
コアメンバ	坂井修一	筑波大学電子・情報工学系助教授 プロセッサ・メモリ混載型 LSI、並列入出力・並列可視化
	朴泰祐	筑波大学電子・情報工学系助教授 並列入出力・並列可視化
	山下義行	筑波大学電子・情報工学系助教授 並列可視化
	和田耕一	筑波大学電子・情報工学系助教授 プロセッサ・メモリ混載型 LSI
	安永守利	筑波大学電子・情報工学系助教授 並列入出力・並列可視化
	星野力	筑波大学構造工学系教授 並列入出力・並列可視化
	白川友紀	筑波大学構造工学系助教授 並列入出力・並列可視化
	中村宏	東京大学先端科学技術研究センター助教授 プロセッサ・メモリ混載型 LSI
	渡瀬芳行	高エネルギー加速器研究機構計算科学センター教授 並列入出力

中澤喜三郎　電気通信大学情報工学科教授
プロセッサ・メモリ混載型 LSI
中田育男　図書館情報大学図書館情報学部教授
並列入出力・並列可視化

2　研究の目的

計算科学の最近の発展は、超並列計算機による計算機の能力の向上と強く結びついている。超並列計算機はベクトル計算機と比較してそのアーキテクチャが多様であり、従って、問題を明確に設定することによって初めて、最先端の半導体技術を駆使した超高性能なシステムを実現できる。この点の好例は、計算科学者の主導により遂行された、先の CP-PACS プロジェクト、GRAPE プロジェクトに見ることができる。CP-PACS プロジェクトは連続体系に焦点をあて、GRAPE プロジェクトは多粒子系にターゲットを絞って、高性能な超並列計算機の開発に成功した。ここで連続体系には、流体力学、格子量子力学、多粒子系には、多体天体系、高分子系などが含まれる。

本報告では「次世代超並列計算機開発」プロジェクトのうちで、連続体向け超並列計算機の開発について述べる[1]。本研究開発の目標は、(1) 高速かつ柔軟な入出力機構・可視化機構・マンマシンインタフェースを実現すること、(2) 超並列計算機の計算速度を現在の 1 TFLOPS から１００倍向上させるための計算機アーキテクチャを考案・検証すること、である。

具体的には、CP-PACS における知見を基に、超高速並列計算機の課題である多量の計算データの並列入出力及び並列可視化のための規範となりうるシステムを実現し、さらに超並列計算機の演算性能そのものの飛躍的増大のために、現在有望と考えられているプロセッサ・メモリ混載型ＬＳＩの開発研究を行う。

3　研究計画の概要

高速な超並列計算機とその結果の解析・可視化を行う外部処理装置の間を柔軟且つ高速に結ぶための並列入出力・並列可視化の研究と、演算性能そのものの高速化を目的とするプロセッサ・メモリ混載型ＬＳＩの開発研究を行う。前者については、計画の前半に、小規模並列システムとグラフィック計算機クラスタを高速スイッチで結合した評価用システムを構築して結合実験・性能評価を行う。計画後半には、この基礎研究の知見に基づいて、超並列計算機と並列入出力・並列可視化装置の結合を行い、これを評価すると共に、ソフトウェアの開発を進めて、実用に耐えるシステムを構築する。プロセッサ・メモリ混載型ＬＳＩの研究開発に関しては、計画前半は、基本方式の検討のためのシミュレータ開発、それを用いてのシミュレーション評価にあて、続いて後半には回路設計・実装実験・実装評価を行う[2]。

3.1　並列入出力・並列可視化システム

3.1.1　研究目的

科学技術計算を主目的とする超並列計算機において、計算における初期データ・中間生成データ・最終結果データの量は膨大であり、この格納及び引き出しを行なうには、大容量・高スループット・低レイテンシの 2 次記憶系が不可欠である。また、このような大規模科学技術計算における演算経過・最終結果の可視化技術は、大量の計算結果の解釈を容易にし、膨大なデータの効率的な保存方法としても、今後一層重要となる。ま

[1] なお、連続体系、多粒子系それぞれの技術開発を最終年度に総合し、柔軟な並列入出力・並列可視化装置を持ち連続体系に適した超並列計算機と多粒子系の専用計算機を組み合わせ、多粒子・連続体混合系に適用可能なヘテロジニアス・マルチコンピュータを構築する。このような統合的システムは、素粒子・宇宙物理学等の学術研究のみならず、地球規模の気候変動シミュレータ、大規模計算力学、物質化学解析、蛋白質立体構造解析等、今後社会的にますます重要度を増すと予想される問題の解決に有用な役割を果たすと期待され、また産業界が開発するコンピュータのプロトタイプともなり得る。

[2] 以上の研究開発を多粒子系の研究開発と統合し、計画最終年度には、連続体系・多粒子系をはじめそれらの混合系にたいしても高い性能を発揮するヘテロジニアス・マルチコンピュータを実現し、素粒子物理学に於ける場の理論シミュレーション、天体シミュレーション、分子動力学計算、粒子的アプローチでの流体計算などに応用する。さらに、フロントエンドとして柔軟なグラフィック処理システム、並列入出力システムと、バックエンドとしての超並列計算機を有機的に結合させる。これによって、高速性と同時に汎用性・柔軟性を持ち、さらにユーザフレンドリな超並列システムが構築されるが、これは現在のところ世界でも他に例を見ないものである。

た、超並列計算機の外部の装置、例えば計算結果に対するデータ整理のような後処理を行なうワークステーションや、上述の可視化装置の運用を超並列計算機と機能分散・並行的に行なうことを想定した場合、超並列機と外部装置とのデータの共有性も重要な要件となる。

以上のような背景を踏まえ、本研究では、超並列計算機に容易に接続可能で、拡張性を持ち、かつ外部環境との柔軟なデータ共有を実現する並列入出力環境及び 2 次記憶系の実現を目的とする。また、その上に並列可視化を支援する環境を構築し、大規模計算と可視化の分散並行処理・リアルタイム処理を実現するシステムを実現する。

3.1.2 要求仕様

並列入出力及び可視化システムに要求される仕様として最も重要なものは、データ転送のスループットである。ここでは、超並列計算機における宇宙物理アプリケーションの例として、流体力学計算における計算時間の見積もりを元に、要求される仕様を算出する。宇宙物理における流体力学計算は、演算時間に対するデータ量が非常に大きくなる典型的な例と考えられ、その要請を満たすデータ転送性能を一つの目標とすることは妥当であろう。

ここでは、リアルタイムな可視化を必要とする、比較的中規模な計算を行なった場合を想定する。格子サイズは 100^3 程度である。ここでは粗目のメッシュにおける比較的短時間の計算において、モデルの振る舞いを概観することを目的と考える。 CP-PACS と同程度の処理装置を用いた場合、 1 step 当たり約 24MB のデータが生成され、超並列機側ではこれを 0.02 秒 /step 程度で生成可能である。このデータを元にスカラ量の等値面図や等高線図、ベクトル量の矢線図などの表示を行なう。ここでは表示データに意味があり、素データを保存する必要はない。

この例では、超並列機のデータ生成に完全に対応する並列入出力・可視化システムに要求されるスループットは約 1.2GB/ 秒であることがわかる。これより、並列入出力チャネルのスループットは全体で少なくとも数 GB/ 秒程度を見積る必要がある。この他、レイテンシの低減に対する要求もあるが、大規模科学技術計算における入出力ではスループットが最優先要件であり、本研究における性能目標もこれを第一と考える。

可視化システム内での画像生成処理に関しては、問題空間のサイズと同時表示させたい変数の種類及び数によって負荷は大幅に変動するため、これを定量的に定めることは難しい。しかし、同時表示したいベクトル線図や濃淡図の要素数が数百万であることを考えると、可視化システムにはフレームバッファを共有する画像処理プロセッサが複数台必要となると考えられる。

3.1.3 実現手法

スケーラビリティを備えた高スループットの並列入出力系及び並列可視化環境を実現するには、単体価格が比較的安価で、超並列システムと入出力システムの間に並列して敷設することが可能な入出力チャネルが必要となる。また、大容量のデータを高速に格納・読出しするための並列ディスク装置及びそれを管理する小規模あるいは中規模の並列処理システムが必要となる。さらに、可視化装置においても、画像生成処理を並列化可能な、フレームバッファ共有型の並列グラフィックシステムが要求される。これらの要素を一つの環境に統合することにより、柔軟で高スループットを持つ並列入出力・可視化システムが実現できると考えられる。

過去の超並列計算機プロジェクトにおいては、多くの場合、入出力系のために専用ハードウェア・ソフトウェアが導入・開発されてきた。しかし、現在の汎用のバスあるいはネットワーク技術は、これらの研究において目標とされてきた水準に十分達しており、今後の超並列計算機用入出力系としては、むしろこういった汎用の製品 (commodity) をハードウェア面で積極的に利用していき、ソフトウェア的な手法でその利用率を高める方向で専用化を進めるのが開発期間・コストに対する効率の点で有利であると考えられる。本研究ではこういった指針に基づき、 commodity ベースのハードウェアを利用していく。

現在、高速入出力媒体として HIPPI や Fiber Channel のように、単体性能で 100 MB/ 秒のオーダの性能を持つものが既に存在する。しかし、これらの入出力媒体はインタフェースが比較的複雑で、単体価格が非常に高価である。本研究で対象とするシステムでは、これに対し 2 桁程度のスループット向上が要求されており、少なくとも数十以上のチャネルを並列に敷設できる必要がある。この点から、より安価で共通性の高い媒体を多数用いることがシステムのスケーラビリティの観点から重要であると考えられる。

一方、 ATM や 100base イーサネットは、単体スループット自体は HIPPI 等より 1 桁低いが、その共用性の高さからインタフェースや媒体・スイッチ等の価格は非常に低くなっている。本研究ではこれらの媒体をベースにスケーラビリティのある入出力チャネルを構成するのが妥当であると考えられる。さらに、数年以内に Gigabit イーサネットのような高速媒体が極めて安価に提供されるようになることは確実と考えられており、研究の最終段階ではこれらの媒体を利用可能であると考えられる。

これらの媒体の中でも、特に 100base-TX によるイーサネットは、既にワークステーションやパーソナルコンピュータにおける実用化が達成されており、特にパーソナルコンピュータ上で運用されているという点から、大量生産に裏づけられたそのコストは他のシステムと比較にならない程度まで低減されている。 commodity 化された高い価格性能比を持つ技術を利用するという、本研究のアプローチから考えると、 100base-TX イーサネットを用いるのが、現段階では最も妥当と考えられる。

3.1.4 システムのイメージ

ハードウェアシステムは以下の 4 つの部分よりなる。

超並列計算機 超並列計算を行なう演算装置である。ここで重要なのは、多数の入出力チャネルを提供するための複数の I/O 専用プロセッサを備えていることである。そして、演算用プロセッサ台数がスケーラブルなだけでなく、この並列 I/O プロセッサ台数にも拡張性が必要である。

並列ワークステーション (並列ディスクサーバ) 多数の大容量ディスク装置と複数の CPU を持ち、並列入出力チャネルを通じて、超並列機の持つ並列 I/O プロセッサとの通信を行なう。必要に応じて、超並列機とのデータ入出力だけでなく、自己でも超並列機と並行に何らかの並列処理を行なうことが可能で、かつ第三者にもファイル共有を可能とするような機能を持たせる。

並列ビジュアリゼーションサーバ 1 つまたは複数の CPU と、強力なグラフィック機能を持つ高速なワークステーション。超並列機および並列ディスクサーバと並列入出力チャネルによって結合され、両者との直接的なデータ交換、あるいはディスクサーバ上のデータ処理を、それらの要素と機能分散的に実行する。

並列入出力チャネル 超並列機・並列ディスクサーバ・並列ビジュアリゼーションサーバの三者を、複数の結合要素で並列結合するチャネルを提供する。高スループットでスケーラビリティのある結合要素が求められる。

これらを有機的に結合し、かつ従来システムより高い効率で機能させるために、以下のようなソフトウェアシステムが必要となる。

アプリケーションレベル 本システムにおける高水準ソフトウェアとして、超並列システムのための効率的なファイル共有システムと、並列可視化アプリケーションを実現することが必要である。前者は NFS のような現行のファイルシステムを見直し、超並列機環境のように、比較的信頼性が高く近距離のネットワークにおいて、ファイル入出力スループットを向上させるものであり、並列ディスクサーバ上で動作させる。後者は超並列機からリアルタイムに生成されるデータをある程度逐次化し、並列度を落としつつ効率的な画像処理を行なうものであり、並列ビジュアリゼーションサーバで動作させる。

ドライバレベル 並列入出力チャネルにおけるデータ転送では、従来の TCP/IP のような汎用・低信頼性ネットワーク用プロトコルに代わる、より効率的な通信プロトコルを用意する必要がある。これは上述のアプリケーションレベルソフトウェアを上位レベルに想定し、必要最低限の機能をもって高スループット・低レイテンシの通信を実現するものである。

3.1.5 関連する研究

いくつかの超並列計算機プロジェクトでは、拡張性を持った並列入出力系の研究が行なわれている。 CP-PACS ではシステム内に最大 128 台の並列 I/O プロセッサが用意され、各々がディスク装置を保持し、大容量・高スループットの入出力系を実現している。しかし、基本的に I/O 系が超並列計算機内で閉じており、外部とのデータ共有や可視化に関する柔軟性に欠ける。

その他の国内プロジェクトとして、 RWCP の RWC-1 及び文部省重点領域研究の JUMP-1 では、各々、専用の入出力系を持っている。 RWC-1 ではプロセッサ間ネットワークとは別にリング状ネットワークと ATM スイッチから成る階層型ネットワークを入出力専用に持ち、画像やセンサー入出力のようなリアルタイム処理を行なうデバイスを結合する。 JUMP-1 では TAXI と呼ばれるバスによって画像出力装置などの周辺デバイスを接続している。

3.1.6　期待される成果

超並列計算の結果生じる膨大なデータを高速に転送・保存・視覚化する技術が今後の大規模科学技術計算において重要な要素技術となることは間違いない。

超並列計算機による大規模科学技術計算、特に連続系のそれにおいては、演算時間に対するデータ量が非常に大きく、シミュレーションを行なう場合もその空間的・時間的範囲は膨大になる。そのため、ある程度粗い精度での予備的演算を行ない、精密な計算を行なう範囲の絞り込みを行なう場合が多い。リアルタイム性を持つ高スループットの並列入出力・可視化システムでは、こういった用途に対し非常に柔軟で効率的なマンマシン・インタフェースを提供することが可能である。

また、高い精度を要求される素データの保存が必要な場合もあるが、系の特徴や時間変化の傾向といったものをアニメーションのような形で保存することは、データ保持効率の点から重要である。本研究で対象とするシステムは、ユーザの要求に応じてリアルタイムな可視化、並列ディスクサーバへのデータの保存、保存されたデータの後処理的な可視化、あるいは外部装置とのデータ共有による機能分散的な並行処理などが可能となる。これらを組み合わせることにより、高い柔軟性と効率を持つ、超並列計算機の周辺環境が実現可能となると考えられる。

さらに、汎用性・共用性の高い入出力媒体をベースとして、ソフトウェア面での効率化を図ることにより、安価で拡張性に富む並列入出力システムを実現する技術を提供できると考えられる。

本研究において構築されるシステムは、複数の並列処理システムの一種の異機種間結合を実現する。例えば、超並列計算機をバックエンドプロセッサに、並列ディスクサーバ及びビジュアリゼーションサーバをフロントエンドプロセッサに、それぞれ機能分担させるような利用法も考えられる。ユーザインタフェースが要求され、かつある程度複雑な処理はフロントエンドで処理し、大規模ではあるが比較的単純な並列性を持つ処理はバックエンドで処理するといった形態が可能である。このように、並列ディスクサーバを単なるディスク管理装置として用いるだけでなく、機能分散型並列処理のための装置として用いることにより、様々なニーズに対応する、柔軟なシステムを構築することも可能であろう。

3.2　プロセッサ・メモリ混載型ＬＳＩ

3.2.1　研究目的と基本方針

本研究の目的は、次世代の連続体向け超並列計算機の要素となる LSI のアーキテクチャを明らかにすることにある。

次節以後で述べるように、 LSI の高集積化・高速化を中心とする計算機技術の発展はめざましいものがあり、これは本研究のターゲットとする２１世紀初頭において持続するものと考えられている。一方で計算科学の要求する計算速度と記憶容量は、２１世紀初頭の計算機技術をもっても最終的な満足が得られるものではない。

そこで本研究においては、まず２００４年前後に実現されるデバイス技術を予測し、これを連続体系の計算科学に適用するための計算機アーキテクチャを考案、シミュレーション評価した後、詳細設計を行い、部分的な試作・評価を行う。

3.2.2　デバイス技術の動向

表 1に米国半導体工業会 (SIA) による今後の LSI 製品（商用マイクロプロセッサチップおよびＤＲＡＭチップ）のロードマップを示す。有名なムーアの法則によれば、半導体の集積度（１チップに納められるトランジスタの数）は、プロセッサ、ＤＲＡＭともに３年で４倍になっている。これは、永続的な法則ではないが、１

９７０年代以後においてほぼ正確に守られており、少なくとも２１世紀の最初の数年間は成り立つものと考えられている。

表 1: SIA による半導体ロードマップ

year	1995	1998	2001	2004	2007	2010
rule (um)	0.35	0.25	0.18	0.13	0.10	0.07
配線レアー数	4-5	5	5-6	6	6-7	7-8
ウェファ直径 (mm)	200	200	300	300	400	400
チップ I/O 数	900	1350	2000	2600	3600	4800
パッケージピン数	512	512	512	512	800	1024
DRAM ビット数	64M	256M	1G	4G	16G	64G
MPU トランジスタ数	12M	28M	64M	150M	350M	800M
MPU チップサイズ (mm^2)	250	300	360	430	520	620
電源電圧	3.3-2.5	2.5-1.8	1.8-1.2	1.5-1.2	1.2 以下	1.2 以下
On-chip Clock Fq. (MHz)	300	450	600	800	1000	1100
I/O Bus Clock Fq. (MHz)	150	200	250	300	375	475
最大消費電力 (W)	12.5	21	36	43	52	62

今、目標とする LSI の実用化の時点を、本研究の終了する３年後、すなわち 2004 年と考える。表１によると、そのときのチップ I/O 数は 2600、 DRAM ビット数 4G、 MPU トランジスタ数 150M、電源電圧 1.5-1.2V、クロック周波数 800MHz、最大消費電力 43W となる。 CP-PACS で使われている技術に比べると、ＤＲＡＭチップの記憶容量で 64 倍、プロセッサのトランジスタ数で約 33 倍、クロック周波数で約 5.3 倍になる。

3.2.3 新プロセッサへの要求仕様

素粒子、宇宙、物性など、計算物理学の主要な分野で要求される計算速度は事実上無限大であると言って良い。実際これらの分野で利用可能な計算速度は、 CP-PACS（2048PU、ピーク性能 614GFLOPS）をはじめ、現在数百 GFLOPS に達しており、これによって大幅な進歩がもたらされているものの、さらに二桁から三桁上の数百 TFLOPS の計算性能の実現によりはじめて真に現実的な計算が可能となる問題も数多くある。素粒子物理においては動的なクォークの効果を完全に採り入れた full QCD 計算と CP 非保存の問題をはじめとする電弱相互作用標準模型への拡張、宇宙物理分野での物質と輻射の相互作用を扱う 6 次元輻射流体問題、物性第一原理計算における数百個の原子の取り扱いなどは、典型的な例である。

これらの計算においては、浮動小数点の計算速度のみならず、この性能に見あった記憶装置から演算装置へのデータの供給性能が要求されることは、特に注意すべき点である。

3.2.4 プロセッサ・メモリ混載型アーキテクチャ

前節までで述べたように、次世代の計算科学向け計算機は、 (1) ピークの演算速度、 (2) 演算装置へのデータ供給速度、の 2 点から最適化されたアーキテクチャをとる必要がある。

- 演算速度の向上

  演算速度の向上は、 (a) クロック速度の向上、 (b) スーパスカラによるプロセッサ内並列化、 (c) 高集積化によって要素プロセッサを多く実装することによる並列化、 (d) 投機的実行による並列化、などによって実現される。

- データ供給系の高速化

  データ供給系の高速化は、メモリ階層のそれぞれにおけるバンド幅およびレインテンシの最適化によって実現される。具体的には、ワーキングセットの大きさとデータアクセスパターンに応じて、メモリバ

スの拡大、メモリの多バンク化、多ポート化、 CP-PACS で開発された疑似ベクタ処理機構の改良、キャッシュの最適化などを行う必要がある。

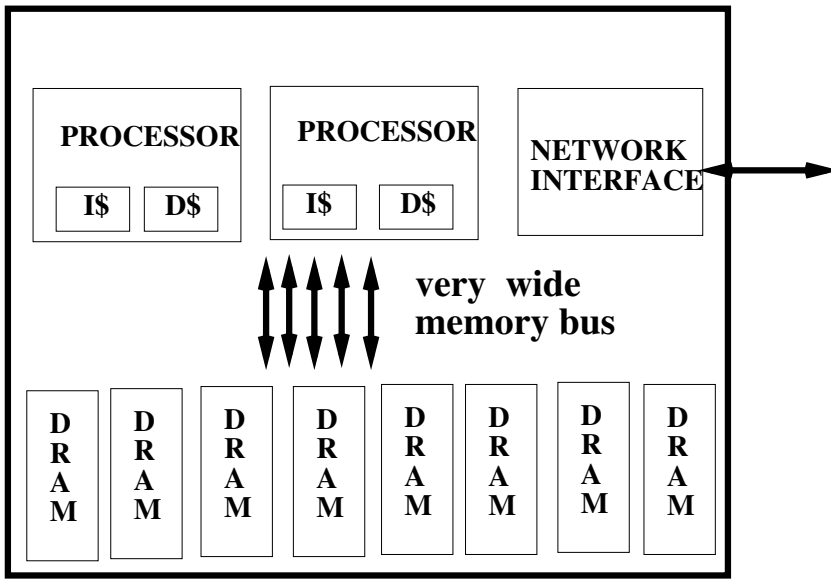
2004 年に実現される高集積チップは、集積度が現在のそれと比較して 33 倍と大きいにもかかわらず、チップの I/O 数はたかだか 3 倍しかない。デザインのバランスを考え直す必要がある。ここで、ＤＲＡＭをプロセッサチップ外部に置く従来のアーキテクチャでは、バスネックになる可能性があり、ワーキングセットの大きくなる科学計算では、その可能性が大きいと考えられる。

この問題を解決するため、プロセッサ・メモリ混載型 LSI の開発を考える。本 LSI においては、メモリバスはチップ内部に実装されるため、バンド幅が大きくレイテンシの小さなバスが実現可能である。具体的には、 DRAM の 1 コラムの幅分のバスを張るなどの方式が考えられる。

プロセッサ・メモリの混載と並んで、同じチップ内にネットワークインタフェースを組み込んで外部との高速通信を実現することも重要な課題である。

- プロセッサ・メモリ混載チップ

  科学技術計算のワーキングセットの大きさは、ほぼ CPU 能力に比例する。したがって、 CPU 速度が N 倍になると、メモリの必要量は N 倍になると考えられる。現在の CP-PACS の PU あたりメモリ量は 64-256MB 程度であり、 2004 年にクロックが 5.8 倍になると、約 400MB-1.5GB が必要となる。これを 1 チップで実装しようとすると、たかだか 1 ないし 2 プロセッサと DRAM の混載チップが適当、ということになる。 PU 間・チップ間のデータ転送を高速化することで、より多くのプロセッサを搭載することが適当となる可能性があり、検討の必要がある (図 1) 。



Processor – DRAM Hybrid Chip

> Clock: 800 MHz
> Parallelism: superscalar (*8) *multi CPU (*2)
> Chip Peak Perf.: 12.8 GFLOPS
> MP system (20K PU) peak perf.: 200~300 TFLOPS

図 1: プロセッサ・メモリ混載型チップ

さらに、プロセッサ・メモリ混載チップにさらに外部メモリを付加する方式、チップ内部に実装するものを (プロセッサと二次キャッシュ) * Ｎとし、 DRAM を外付け (実際には MCM を利用) するなどの方式を検討する。

- 期待される性能

  2004 年の計算機は、 CP-PACS と比較して、クロック周波数で 5.3 倍 (800MHz)、チップ内並列性能で 8 倍、実装規模で約 10 倍となることが期待され、単純計算によると、ピーク性能は 200TFLOPS-300TFLOPS となる。前述したようにこの場合、メモリバンド幅、キャッシュ速度、レジスタバスのバンド幅のそれぞれで、これに見合う速度のデータ供給がなされなければならない。

3.2.5 関連する研究

プロセッサメモリ混載型のチップアーキテクチャの研究開発は米国および日本を中心に盛んに進められている。中でも、ＵＣＢの IRAM、ウィスコンシン大学の multi-scalar、九州大学の PPRAM などが有名である。

これらは、本来、汎用チップをめざしたものである。本研究開発は、これらに対してワーキングセットが大きく計算の規則性が高い計算科学をその応用と想定しており、メモリバンド幅の拡大やメモリパイプラインの強化など、大規模のデータを供給するためのアーキテクチャに重点が置かれるのが特徴的である。

3.2.6 期待される成果

数百 TFLOPS の計算性能が実現された場合、素粒子物理においては、充分な規模の full QCD 計算が可能となり、これによって強い相互作用全般にわたり QCD 第一原理に基づく検証と予言が現実のものとなるのみならず、電弱相互作用における CP 非保存の問題をはじめ、素粒子標準模型全体とその拡張に関わる研究の躍進が期待される。また、物質と輻射の相互作用を忠実に採り入れた 6 次元宇宙輻射流体力学計算、数百個の原子を対象にして、物質の基底状態と励起状態を量子力学に基づいて現実的に解き明かす物性第一原理計算など、現在のところ実現不可能なスケールの諸計算が可能となる。即ち、数百 TFLOPS の計算性能の実現は、ミクロからマクロにわたる物質の諸相の予測と解明を、真に現実的な計算パラメータ下で可能にするという意味で、計算物理学に新たな時代を劃するものと期待される。

4 平成 9 年度の研究成果の概要

今年度の主な成果は以下のとおりである。

- 実験用 CP-PACS システムに高速データ交換インタフェースを搭載し、これを高速スイッチを通して入出力処理システムに接続した評価用システムを構成し、これを用いて並列入出力の基礎的研究を行なった。
- プロセッサ・メモリ混載型 LSI の基本方式の検討をシミュレータ開発を通して行なった。

以下、それぞれについて簡単に述べる。

4.1 並列入出力・並列可視化システム

今年度の目標は、現行の commodity ハードウェア及びソフトウェア技術の性能的限界を調査することにあった。このため、まず当面の並列入出力チャネルのターゲットである 100base-TX イーサネット上で、データ生成系である並列計算機と並列ディスクサーバである並列ワークステーションを 100base-TX 対応のスイッチで結合したシステムを構築した。並列ディスクサーバとしては、次年度以降の各種機能分散処理実験への応用も視野に入れ、共有メモリ型の並列ワークステーションを導入したものである。

具体的には、並列データ生成系として超並列計算機 CP-PACS の小規模システム (CP-PACS サブシステムと呼ぶ) を、並列ワークステーションとして SGI/Cray Origin-2000 システムを設置し、これらを 100base-TX イーサネット及びそれに対応するイーサネットスイッチ (Baystack 350T) で結合した。当面の目的を勘案し、双方の計算機に各 4 チャネルずつの入出力ポートを設けた。スイッチはこれらの通信容量を十分満たすものであるため、ピークでは 400 Mbps の単方向通信が可能なシステムとなっている。 CP-PACS サブシステム側は各イーサネットポート毎に IOU(入出力プロセッサ) が用意されており、データ転送スループット的に問題はない。 Origin-2000 側は、 4 チャネルの入力を各々処理し、なおその他のサービスを行なえるよう、 8 台のプロセッサを用意している。

今後、これらのシステム上で、様々な状況の下、 NFS を始めとする既存の各種データ共有 / データ転送ソフトウェアがどのような性能・振舞いを見せるかを詳細に調査する。その結果を基に、単体チャネル、チャネル本数、ディスクサーバ台数等に関し、本研究で想定すべき数値目標について検討する。最終的に、今後数年間で実用化が期待される技術における推定性能との比較により、システム全体のスケール予想を立てる。

並列可視化システムに関する研究としては、今年度は特別なハードウェア装置の導入は行なわず、現在の並列入出力システムを中心としたデータ供給性能が、次年度以降に導入予定の表示系に対してどの程度満足できるものであるかについて調査した。

平成 9 年度実験計画で構築したシステムのイメージを図 2に示す。

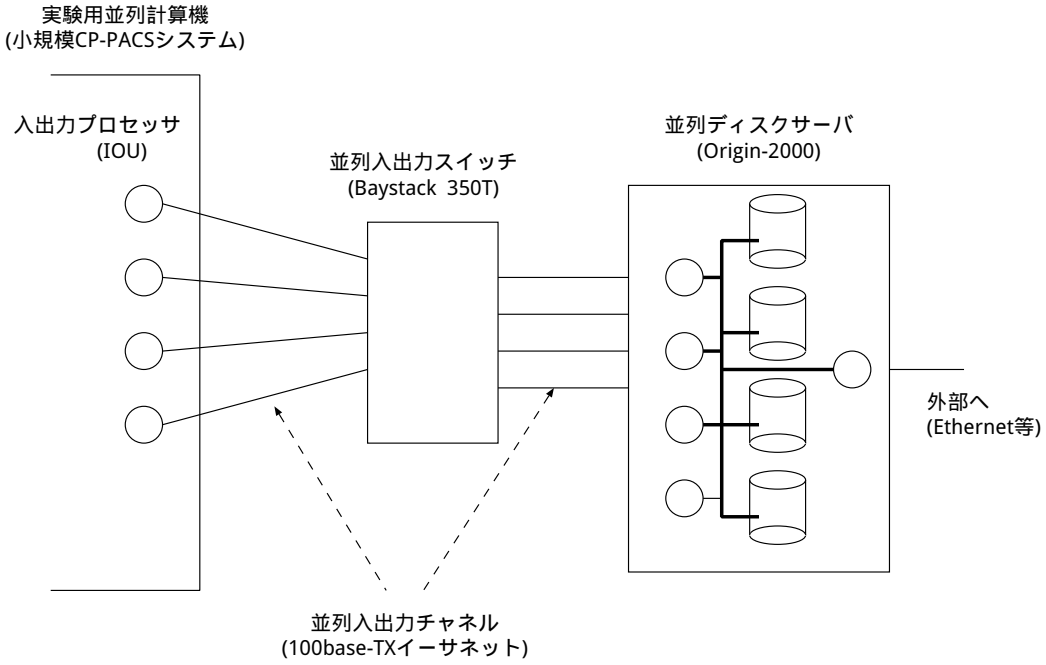


図 2: 平成 9 年度 並列入出力・可視化システム実験装置イメージ

以上、平成 9 年度は現行技術による性能限界を測定し、各部の増強のファクタを整理した。この結果、ハードウェア・ソフトウェアの両面において、各々の目標が明らかになった。これを基に、実用システムにおいて実現可能と思われる単体チャネル性能とその並列度を求めた。さらに、現行ソフトウェア (アプリケーションレベル及びデバイスドライバレベル) において重点的に改良すべき点を明らかにした。

4.2 プロセッサ・メモリ混載型 LSI

平成 9 年度は、プロセッサ・メモリ混載型 LSI の基本方式の検討をシミュレータ開発を通して行なった。前提となっているのは、 2004 年のデバイス技術である。

シミュレータは、以下のように構築している。

- Stanford 大学で開発されたシミュレータ Sim OS を移植し、今回の研究目的である連続体系の計算科学向けのベンチマークによってこれを評価する。その際、ミュレーション速度とデバグ・性能評価時の視覚化、さらに移植性を考慮し、汎用ワークステーションをメモリ拡張するなどして用いる。具体的には、今年度購入した SGI Origin-200 をこの用途に用いている。
- Sim OS でアーキテクチャパラメータを様々に変えて評価することで、各パラメータに対する性能の感度 (sensitivity) を調べる。

- Sim OS を一部改造する形で、アーキテクチャを最適化する。そのさい、 CP-PACS のプロセッサである HARP1E のシミュレータを参考にし、場合によっては一部を組み込む。

- 汎用性と高速性を考慮して、 C 言語を用いたシミュレータを書くことを検討する。

アーキテクチャの検討事項は、 (1) 実装規模、動作速度、 (2) メモリアクセスの方式、 (3) バス幅の拡大、多ポート化、多バンク化、 (4) 混載する CPU の数 (5) CPU へのデータ供給系をパイプライン化する工夫、 (6) キャッシュの方式。特にコヒーレンス制御の方式、 (7) スケジューリング方式、 (8) チップ外通信のためのアーキテクチャなどである。

これらのうち、主に最初の 3 点についての評価・検討が進められている。具体的には、 2004 年の半導体技術を前提とした場合、数多くのＣＰＵを１チップに搭載したＬＳＩアーキテクチャよりも、 2 ギガビット程度のメモリと少数（１ないし４個程度）のＣＰＵを搭載したアーキテクチャを採用し、ＣＰＵ・メモリ間のバスのバンド幅を大きくする混載型アーキテクチャの優位性が、大規模科学技術計算に対しては確認されつつある。より定量的な評価と設計の詳細化は次年度以後に行われる。

5　今後の計画

以下に、来年度以降の計画について簡単に述べる。

平成 10 年度

並列入出力評価用の入出力装置を拡張するとともに、グラフィック計算機クラスタに接続し、並列入出力と並んで並列可視化の研究開発を行なう。

プロセッサ・メモリ混載型 LSI については、前年度に検討した基本方式に対してシミュレーション評価を行ない、検討・改善を加える。また、基本方式に則った回路設計の準備を開始する。

平成 11 年度

並列入出力・並列可視化のための結合システムの評価を続行し、素粒子物理・宇宙物理・物性物理等の実用計算に試行することにより、実用化のための改善項目整理を行なう。これを基に実用的システムの構築のためのスイッチ拡充とグラフィック処理システムの強化を行なう。

プロセッサ・メモリ混載型 LSI については、前年度迄のシミュレーションによる性能評価に基づいて、 LSI の回路設計を本格化する。

平成 12 年度

前年度までの小規模並列システムでの結合システムにおける知見をもとに、超並列計算機と並列入出力・並列可視化装置の結合を行ない、これを評価するとともに、素粒子物理・宇宙物理・物性物理等の実用計算への応用を開始し、また実用化のためのソフトウェア開発に着手する。

プロセッサ・メモリ混載型 LSI については、回路設計とシミュレーションによる論理検証を行ない、 LSI の実装実験を進める。

平成 13 年度

超並列計算機と並列入出力・並列可視化装置の結合システムの最終評価を行ない、実用化のためのソフトウェア開発を継続して、超並列計算機での素粒子物理・宇宙物理・物性物理等の結果の解析に応用する。

プロセッサ混載型 LSI の実装・評価を行なう。

多粒子系システムと結合し、ヘテロジニアス・マルチコンピュータシステムを構築する。応用としては、分子動力学および流体計算への適用を図る。

計画最終年度にあたり、本計画による技術開発の整理・評価を行い、次のステップの方向性を探るための会議を開催する。

6 学会発表、論文等

6.1 国際学会等における発表

[1] Y. Iwasaki, *Status of the CP-PACS Project*, in Proceedings of Lattice '96 (St. Louis, USA, 4-8 June, 1996), Nucl. Phys. B (Proc. Suppl.) 53 (1997) 1007-1009.

[2] Y. Iwasaki, *The CP-PACS Parallel Computer Project*, in Proceedings of International Conference "Multi-Scale Phenomena and Their Simulation", eds. F. Karsch, B. Monien and H. Satz, World Scientific (1997) 80-90.

[3] Y. Iwasaki, *The CP-PACS Project and Computational Physics*, in Proceedings of International Symposium on Parallell Computing in Engineering and Science, Science and Technology Agency (1997), to appear.

[4] Y. Iwasaki, *The CP-PACS project*, in Proceedings of the International Workshop "Lattice QCD on Parallel Computers" (Tsukuba, March 10-15, 1997), Nucl. Phys. B (Proc. Suppl.), to appear.

[5] A. Ukawa, *The CP-PACS Parallel Computer*, in Proceedings of CHEP'97 (Berlin, April 7-11, 1997) 595-600.

[6] CP-PACS Collaboration (S. Aoki *et al.,*), *CP-PACS results for quenched QCD spectrum with the Wilson action*, in Proceedings of International Workshop "Lattice QCD on Parallel Computers" (Tsukuba, March 10–15, 1997), Nucl. Phys. (Proc. Suppl.), to appear.

[7] CP-PACS Collaboration (S. Aoki *et al.,*), *Full QCD results from CP-PACS*, in Proceedings of International Workshop "Lattice QCD on Parallel Computers" (Tsukuba, March 10–15, 1997), Nucl. Phys. (Proc. Suppl.), to appear.

[8] CP-PACS Collaboration (S. Aoki *et al.,*), *CP-PACS results for the quenched light hadron spectrum*, in Proceedings of Lattice '97 (Edinburgh, Scotland, 22-26 July, 1997), Nucl. Phys. (Proc. Suppl.), to appear.

[9] CP-PACS Collaboration (S. Aoki *et al.,*), *Hadron spectroscopy and static quark potential in full QCD: A comparison of improved actions on the CP-PACS*, in Proceedings of Lattice '97 (Edinburgh, Scotland, 22-26 July, 1997), Nucl. Phys. (Proc. Suppl.), to appear.

[10] T. Nakamoto, M. Umemura, and H. Susa, *Photoionization of a Clumpy Universe*, International Symposium on Supercomputing, *New Horizon of Computational Science* (1997) in press

[11] Y. Abei, K. Itakura, T. Boku, H. Nakamura, K. Nakazawa, *Performance Improvement for Matrix Calculation on CP-PACS Node Processor*, Proc. of HPC Asia'97 (Seoul, Korea, Apr. 1997), 672-677.

[12] K. Itakura, T. Boku, H. Nakamura, K. Nakazawa, *Performance evaluation of CP-PACS on CG benchmark*, in Proc. of HPC Asia'97 (Seoul, Korea, Apr. 1997), 678-683.

[13] T. Boku, K. Itakura, H. Nakamura, K. Nakazawa, *CP-PACS: A massively parallel processor for large scale scientific calculations*, in Proc. of ACM Int. Conf. on Supercomputing'97 (Vienna, Austria, Jul. 1997), 108-115.

[14] S. Sakai, *Seamless Computing: Integrating Parallel Computing and Network Computing with Future PCs*, HiPC'97, Bangalore, India, December 18, 1997.

[15] A Data Alignment Technique for Improving Cache Performance Preeti Ranjan Panda, Hiroshi Nakamura, Nikil D. Dutt, Alexandru Nicolau, International Conference on Computer Design (ICCD '97), Austin, October 1997

[16] Improving Cache Performance through Tiling and Data Alignment Preeti Ranjan Panda, Hiroshi Nakamura, Nikil D. Dutt, Alexandru Nicolau 4th International Symposium on Solving Irregularly Structured Problems in Parallel (IRREGULAR'97), Paderborn, June 1997

[17] K. Sayano and T. Shirakawa *Pleiades: A Prototype of Inter-processor Network Generation System*, Proceedings of the 1997 International Symposium on Parallel Architectures, Algorithms and Networks (I-SPAN'97), Taipei, Taiwan, Dec. 18-20, 1997, pp.202-206.

6.2 国内学会等における発表

[1] 板倉 憲一, 松原 正純, 朴 泰祐, 中村 宏, 中澤 喜三郎, 超並列計算機 CP-PACS における NPB Kernel CG の評価, 並列処理シンポジウム JSPP'97 論文集 (神戸, 1997 年 5 月), 5-12.

[2] 服部 正樹, 松原 正純, 板倉 憲一, 朴 泰祐, 超並列計算機 CP-PACS における分子動力学法シミュレーション, 情報処理学会研究報告 97-HPC-66-2 (1997), 7-12.

[3] 松原 正純, 板倉 憲一, 朴 泰祐, 中村 宏, 中澤 喜三郎, 超並列計算機 CP-PACS のネットワーク性能評価, 情報処理学会研究報告 97-HPC-67-10 (1997), 55-60.

[4] 松原 正純, 板倉 憲一, 朴 泰祐, 超並列計算機 CP-PACS における空間分割法による分子動力学法シミュレーション, 情報処理学会研究報告 97-HPC-69-10 (1997), 55-60.

6.3 雑誌論文等

[1] Y. Iwasaki, K. Kanaya, S. Kaya, and T. Yoshié, *Scaling of chiral order parameter in two-flavor QCD*, Phys. Rev. Lett. 78 (1997) 179-182.

[2] Y. Iwasaki, K. Kanaya, T. Kaneko, and T. Yoshié, *Scaling in SU(3) pure gauge theory with a renormalization group improved action*, Phys. Rev. D56 (1997) 151-160.