

(I-28) PCクラスターを用いた自作並列計算機の構築とその有効性の検討

中央大学 学生員 ○須江 克章
中央大学大学院 学生員 山口 敏
中央大学 正会員 檜山 和男

1. はじめに

近年、偏微分方程式の大規模数値計算に対して計算時間の短縮と記憶容量の分散を可能にする並列計算が広く行われるようになってきている。しかし、現在の並列計算環境は並列専用計算機がまだまだ高価なものであるため、利用者数や使用時間が制限されている状況にあり、手軽に並列計算を行うことができない。

そこで本報告では、PC(Personal Computer)8台をEthernetで接続し、通信ライブラリにMPI(Message Passing Interface)を搭載した仮想並列計算機¹⁾を構築した。そして、領域分割を使用した並列有限要素法を用いて自作並列計算機の性能評価、特にHUBの性能の差異が並列計算効率に及ぼす影響について検討を行った。

2. 並列計算機システム

並列計算機システムはメモリの割り当て方から2つの種類に分類できる。

- 共有メモリ型並列計算機
- 分散メモリ型並列計算機

共有メモリ型並列計算機は複数のCPUが1つのメモリを共有するシステムであり、データ分散を意識したプログラミングを行う必要がない。しかし、全プロセッサのアクセスが1つのメモリに集中するため、プロセッサ数の拡張には限界がある。一方、分散メモリ型並列計算機はCPU-メモリから構成されるノードをネットワークで接続したものであり、ノードを増ややすく拡張性が高い。しかし、プロセッサは同じノード上のメモリしかアクセスできないため、他のノード上にあるデータを利用するためにはノード間でメッセージを介して通信する必要がある。そのため、MPIやPVM(Parallel Virtual Machine)などの通信ライブラリを用いて、分散データを意識したプログラミングを行う必要がある。また、ノード間のネットワーク通信がボトルネックとなり十分な並列化効率を得られない場合がある。

自作並列計算機は各ノードをPCに置き換えることで仮想的に並列計算機として機能させるものであり、分散メモリ型並列計算機の種類に入る。PCを用いることでコストパフォーマンスが高く、構成に柔軟性があるシステムを構築することができる。通信ライブラリについてもOS(Operating System)は、PCで使用できるPC-UNIXからWindowsまでサポートしている。

3. 自作並列計算機の構成

a) PC構成

各PCにはFUJITSU社製FMV-590TA2を用いた。PCの構成を表-1に示す。ここに示す部品は、より速いCPU、大容量のメモリに変更することも可能である。

表-1 PC仕様

部品	仕様
CPU	Pentium 90MHz
Cache	256KB
RAM	48MB (8MB SIMM×6)
NIC	DEC DC21x4x PCI(10/100Base-Tx)

次に使用したソフトウェアについて説明する。一覧を表-2に示す。OSには安定性、性能、コストの面からPC-UNIXの一種であるLinux²⁾を選択した。通信ライブラリにはLinuxで動作するMPICHを用いた。また、ネットワークシステムとしてNFS(Network File System)を採用したことで、PC台数を意識することなくプログラムの編集、コンパイルを行うことができる。これら使用したソフトウェアは全てGNUライセンスに基づくフリーウェアであるため、ソフトウェアについてはコストをかけることなく構築することができる。

表-2 ソフトウェア一覧

OS	Linux-2.0.34
コンパイラ	g77,gcc
通信ライブラリ	MPICH1.1.1
ネットワークシステム	NFS

b) ネットワーク構成

PC8台をEthernetで接続し、その内の1台をGateway PCとすることにより、他のネットワークへの接続を実現している。PC間のネットワーク環境を比較するためにHUBには、10Base-TのHubと100Base-TXのHubと100Base-TXのスイッチングHUBの三種類を用意した。全体の構成図を図-1に示す。

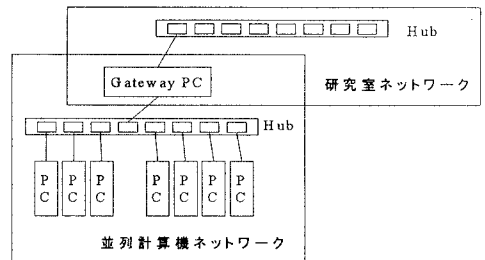


図-1 ネットワーク構成図

Key Words: 自作並列計算機, MPI, 有限要素法
〒112-8551 東京都文京区春日1-13-27
TEL. 03-3817-1815 , FAX. 03-3817-1803

4. 数値解析例

数値解析例として、安定化有限要素法³⁾による非圧縮粘性2次元円柱周り流れの解析を行った。有限要素分割には、節点数20,616、要素数20,334を用いた。図-2にFarhatの領域分割法⁴⁾により8分割した領域分割図と境界条件を示す。微小時間増分量を0.001秒とし、ステップ数を200回とした。(並列処理の詳細については参考文献5)を参照)

ネットワーク環境性能(HUBの差異)とプロセッサ数を変化させた時のCPU時間と通信時間の内訳を表-3に示す。表中において各プロセッサの通信時間のばらつきを示すため、通信時間については最大、平均、最小の値を示している。これより通信時間にかなりのばらつき(平均値に対して4~17%程度)あることがわかる。この原因は領域分割において、各プロセッサの通信量の均等化が図れていないことによる。また、10Base-T環境ではプロセッサ数を増やすことで100Base-TX環境と比べて効率が大幅に低下していることがわかる。その理由として、100Base-TX環境では伝送速度は100Mbps(理想速度)であるのに対し、10Base-T環境では伝送速度が10Mbps(理想速度)しかないため、プロセッサ数が増えるにつれてデータの衝突の頻度が増え、効率が低下したと考えられる。次に100Base-TX環境でのスイッチングHUBの効果について考える。スイッチングHUBとは出力ポートが異なる場合に複数の入力をデータの衝突なしに同時に転送できる機能を有しているHUBのことで、ネットワークが大規模化した際に有効である。結果として、今回構築した自作並列計算機では使用できるプロセッサ数が8台までだったため、100Base-TXのスイッチング機能なしのHUBとの差異が見られなかった。図-3に演算速度倍率を、図-4に並列化効率を示す。今回の数値解析例は比較的小規模な問題であったため、計算時間に比べて通信時間の占める割合が大きく、高い演算速度倍率と並列化効率を得ることが出来なかった。

5. おわりに

本報告ではPCクラスターによる並列計算機を構築し領域分割を使用した並列有限要素法を用いて性能評価、特にHUBの性能の差異が並列計算効率に及ぼす影響について比較検討し、以下の結論を得た。

- 1. 10Base-Tと100Base-TX環境では大きな性能の差異が見られ、PC間のネットワークには100Base-TX環境を整える必要があるといえる。
- 2. PCが8台程度の場合にはスイッチングHUBとスイッチング機能を持たないHUBとの差異は見られなかった。

今後の課題として、大規模な数値解析を行い本計算機システムの有効性を示す。またプロセッサ数を増やしより大規模な並列計算機を構築し、その場合の並列計算効率を調べるとともに、性能の異なるPCが混在する場合の領域分割法を検討する。

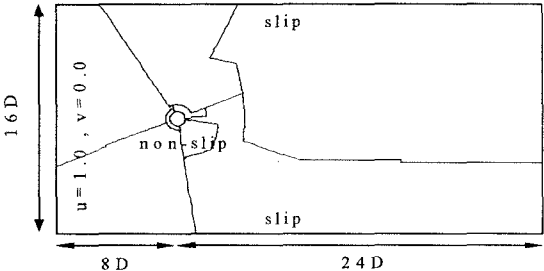


図-2 領域分割図(8CPU使用時)と境界条件

表-3 CPU時間と通信時間の内訳

PE	100Base-TX Switching Hub		100Base-TX Hub		10Base-T Hub	
	全 CPU	通信時間	全 CPU	通信時間	全 CPU	通信時間
1	3622.8	0.0	3622.8	0.0	3622.8	0.0
4	1438.8	最大	1438.0	368.0	1861.7	788.4
		平均		343.5 (23.9%)		746.9 (40.1%)
		最小		294.3		709.5
8	1023.6	最大	1035.7	410.5	2769.8	2030.1
		平均		383.2 (37.0%)		2021.6 (72.9%)
		最小		343.3		2009.1

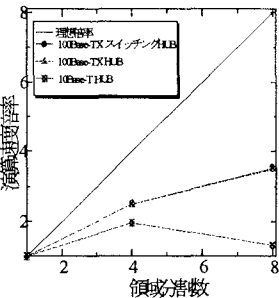


図-3 演算速度倍率図

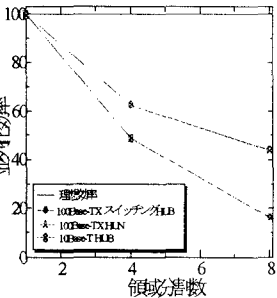


図-4 並列化効率図

参考文献

1) 塩谷隆二, 矢元基, "自作並列計算機とその計算工学への応用" 計算工学講演会論文集 Vol.2, pp155-158, 1997.
2) はねひでや, やまだあきら, あべひろのぶ, "るんるんLinux": アスキー出版局.
3) 櫻山和男, 猪股渉, "混合補間要素を用いた非圧縮粘性流れ解析のための高精度安定化有限要素法", 土木学会論文集 No.591/I-43, pp125-137, (1998).
4) Farhat, C.: "A simple and efficient automatic FEM domain decomposer", Computers & Structures, 28, pp576-602, 1988.
5) 玉井典, 櫻山和男, "非構造格子に基づく三次元非圧縮粘性流れの並列有限要素解析" 第11回数値流体力学シンポジウム論文集, pp557-558, 1997.