# 並列計算機へのデータマッピング

RA 蓬来祐一郎
情報理工学系研究科コンピュータ科学専攻

## 概要

多くの科学技術計算ではシミュレーションを始めとして反復計算を行う．このような並列計算では，データの依存関係を考慮し適切に各計算ノードに仕事を割り当てることができれば，通信が削減され，安定した並列性能が得られる．この研究ではヘテロな環境も含め，適切なデータのマッピングを行える手法の開発を目指す．

## 1 はじめに

多くの科学技術計算において，並列化は欠かせない要素となりつつある．一方，並列計算機は多様化しており，ネットワーク性能，ネットワークトポロジーも多様である．このような環境では，均一性を仮定した既存手法の多くは負荷の不均衡を起こす可能性がある．

計算に必要なデータの依存関係がわかっている場合には，それをグラフを用いて表現し，得られたグラフを分割する手法がしばしば用いられる．グラフの各頂点に計算負荷，各枝に通信量を表す重みを与え，目的のプロセッサ数にグラフの頂点を分割するのである．このとき，各パーティションに割り当てられた頂点の重みを均等にし，異なるパーティションをまたがる枝の重み(枝カット)が最小になるようにする．このような分割を行うことにより，計算負荷が均等になり，データの局所性が抽出され通信が削減された結果として，高度な並列性能が期待できる．

しかし，従来手法で並列計算のデータ分割を行う場合に，3 つの欠点がある．1 つ目は，枝カットは総通信量を正しく反映していない点．2 つ目は，通信時間は，総通信量を抑えるだけでなく個々のプロセスの通信量を抑えることが重要である点．3 つ目は，通信が必要なパーティションは，並列計算機上で近いノードに割り当てられることが望ましいが，これが考慮されない点である．そこで，これらの問題点をカバーする手法について研究を行った．

既存研究から，データを再帰的にマルチレベルな二分割を行う手法は，非常に有効な手法であることが知られている．本研究では，この手法を拡張し，二分木で並列計算機を階層的に表現することによって，再帰的な二分割を行う．

## 2 計算グラフ

まず，通信量の正確な見積もりのために，計算グラフを定義した(図 1)．計算グラフでは，枝に通信するべきデータが関連付けられている．ハイパーグラフにより総通信量を正確に見積もる研究もあるが，通信の方向性，通信データの表現力の点で並列計算におけるモデルとして適している．

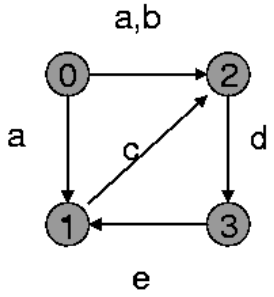


図 1　計算グラフ

## 3 ネットワークモデル

通信時間は遅延とバンド幅でモデル化されることが多いが，通信遅延の一部は，ほかの通信で隠蔽でき，また通信が多い場合に重要となるのがバンド幅であるため，各資源のバンド幅で通信のモデル化を行った．通信資源をもとに並列計算機は，階層的に表現される(図 2)．これをもとに，グラフは再帰的に二分割され，目的の数に分割する(図 3)．縮約されたノード間の通信は近似される．これにより，ネットワーク構造に適した分割

が可能になる．

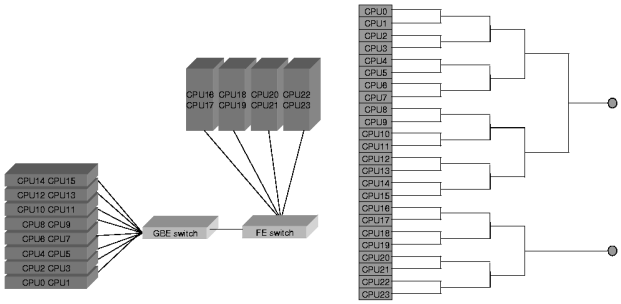


図 2　並列計算機の階層化

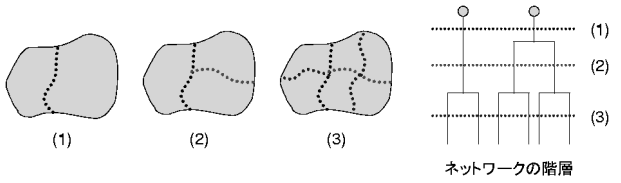


図 3　Recursive Bisection

## 4 通信コストの最小化

総通信量を抑えるのではなく，通信時間を抑えるには，個々の資源によって伝送されるデータ量を抑える必要がある．この点を考慮し，Kernighan-Lin Fiduccia-Mattheyses アルゴリズムを拡張し，通信コストを最小化する．これにより，個々のプロセスの通信量を均衡化させるだけでなく，通信が集中するような資源を用いる通信を抑制することが期待できる．

## 5 分子動力学法での実証実験

提案手法を生体分子の分子動力学法プログラム MolTreC2-DM に実装した．従来の MolTreC2 は，全粒子のデータを全プロセスが共有する粒子分割法を用いており，各ステップで位置情報の更新に集団通信の all-gather を用いている．それに対し, MolTreC2-DM では, 空間を小セルに分け，個々のプロセスが直接に必要なデータの通信を行う．図 4 は提案手法による分割の例である．図 5 に他の分割手法と合わせ，分子動力学法を図 2 のヘテロクラスタで実行した場合の，各反復における通信時間を示す．2 つのクラスタをつなぐ回線に通信が集中するが，提案手法では改善され，通信時間が短くなっている．

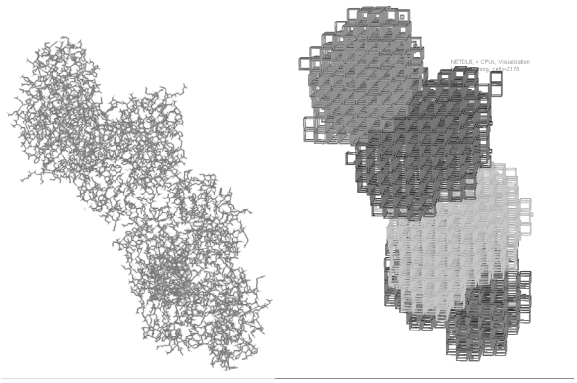

図 4　PDB-1XS2 とその 4 分割の例

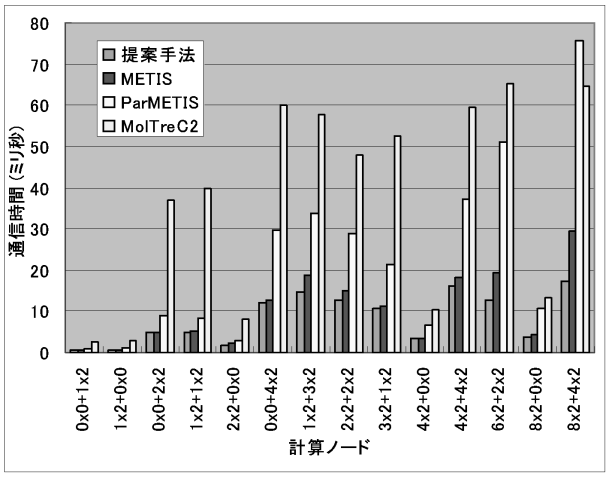


図 5　ヘテロなクラスタでの通信時間

## 6 おわりに

既存のグラフ分割手法の問題点を改善し，並列計算に適したグラフのマッピング手法を開発した．提案手法により，計算負荷を均等に保ちながら通信時間を大きく削減できることを示した．この手法により，ロバストな並列性能が期待でき，また，多くの科学技術計算が同様の計算パターンを持つため，多くのアプリケーションで同様の効果が期待できる．