固定分配セルを用いた動的負荷分散法による並列分子動力学法シミュレーション

林 亮子[†] 堀口 進[†]

並列計算機上で行う分子動力学法シミュレーションは，材料設計の重要な手段の1つとして期待されている．短距離相互作用が働く分子動力学法の有効な並列化手法の1つは，シミュレーションを行う空間を分割し，プロセッサ要素に割り当てるドメイン分割法である．しかしドメイン分割法では，計算負荷の不均衡が並列処理効率を低下させる．そこで本論文は，ドメイン分割法で用いるセルを単位とした計算負荷移動による動的負荷分散法を提案する．本論文では，プロセッサ要素間の通信パターンを保存しながら計算負荷を移動するために，ドメインの中で固定されたセルである固定分配セルを導入する．この動的負荷分散法を並列計算機 CM-5 上に実装し，計算負荷の平均化性能について詳しく検討し，その有効性を示す．固定分配セルは計算負荷の移動を制限するため，固定分配セル数と平均化可能な計算負荷の集中度との関係を定式化し，議論する．

A Parallel Molecular Dynamics Simulation by Dynamic Load Balancing Based on Permanent Cells

RYOKO HAYASHI[†] and SUSUMU HORIGUCHI[†]

Parallel Molecular Dynamics simulations have been widely noticed as an important basic technique for material design. Domain Decomposition Method is one of efficient parallel methods for short-range classical Molecular Dynamics simulations. However, Domain Decomposition Method has a problem of computational load unbalancing. To remedy this problem, we propose a dynamic load balancing method based on cell redistribution. We introduce a concept of permanent cells in order to reduce the communication overhead between PEs. The performances of the proposed dynamic load balancing method are discussed by implementing it on a parallel computer, CM-5. It is seen that the proposed method achieves sufficient load balancing for each PE and reduces the execution time of parallel Molecular Dynamics computations. Moreover, we analyze the relation between the concentration of molecules and the effective region of computational load balancing theoretically.

1. はじめに

自然界の物質は，微視的には分子の集団である．分子動力学法（MD）とは，物質を構成する分子や原子1つ1つの運動を模擬することによって物質の状態の時間変化を調べるシミュレーション手法である[1]．MDでは各分子の位置と運動量の時間発展を調べることができ，材料科学の分野において物質・材料設計の有力な手段として注目されている．そして，より複雑な物質や詳細な現象を扱うために，並列処理による MD の高速化について多くの研究が行われている[2]．

ドメイン分割法は，相互作用の影響範囲が小さい短距離 MD の，効果的な並列化手法の1つである[2)~4]．ドメイン分割法では，シミュレーション空間を相互作用の影響範囲の大きさの小空間（セル）に分割し，セルをまとめたドメインを並列計算機のプロセッサ要素（PE）に割り当てる．ドメイン分割法は，セルを使用することによって計算量を削減し，ドメインによって PE 間で共有するデータを小さくして通信時間を削減する．しかし，ドメイン分割法はシミュレーション空間を PE に割り当てることから，分子がシミュレーション空間中に不均一に分布する場合には，各 PE の計算負荷に不均衡が生じ並列化効率が著しく低下する．分子の集中する位置は予測不可能であり，計算負荷が集中する PE も予測不可能である．そのため，並列処理における計算負荷の平均化には，動的に計算負荷を移動する動的負荷分散法が必要である．

2次元空間の MD シミュレーションでは，シミュレーション空間を分割する並列化手法において，分割

† 北陸先端科学技術大学院大学情報科学研究科
School of Information Science, Japan Advanced Institute of Science and Technology

領域の境界移動により計算負荷を移動する動的負荷分散法が提案されている．Brugé と Fornili[5] は 2 次元空間を短冊状に分割して PE に割り当て，分子の数を均等にするように連続的に境界を移動させる方法を提案し，計算負荷を平均化した．Kohring[6] は 2 次元空間のセルを用いたドメイン分割法による粒子シミュレーションに，セルの列単位で計算負荷を移動する動的負荷分散法を導入して計算負荷の平均化を行った．

一方，3 次元空間の MD シミュレーションでは多数の PE を使用する場合が多く，一般に PE 間の近傍関係がより複雑である．そのため，PE 間の近傍関係を保存したまま計算負荷を移動することが困難である．計算負荷移動の方向性および移動量に制限を加えない場合は，PE の近傍関係が壊れ，PE 間通信時間が増大する．PE の近傍関係を保存したままドメイン境界を移動するためには，シミュレーション空間の 1 つの断面全体を移動する必要がある．このとき，境界移動の影響が複数のドメインに及ぶことと，移動する計算負荷量が大きくなりすぎることが問題である．したがって，現在までの 3 次元空間の MD シミュレーションでは，シミュレーション空間を 1 つの次元で分割可能な場合以外は，動的負荷分散に関する研究は十分に行われていない．

本論文では，ドメイン分割法を用いた 3 次元 MD シミュレーションに適用するための動的負荷分散法（Dynamic Load Balancing: DLB）を提案する．PE 間の近傍関係を保存するために「固定分配セル」を導入した DLB は，ドメイン分割法のセルを計算負荷の単位として使用する[4),7),8)]．固定分配セルは移動する計算負荷の量，移動方向に制限があるため，並列計算機 CM-5[9] 上での MD シミュレーションの実装により，DLB の有効性を検討する．さらに固定分配セル数と計算負荷の集中度との関係を理論的に検討し，実装結果と比較してその妥当性を議論する．

本論文の構成は以下のとおりである．2 章では MD シミュレーションおよび並列化手法ドメイン分割法の概要を示す．3 章では DLB の概要を示し，計算負荷移動の手順について述べる．4 章では DLB を使用した DLB ドメイン分割法による並列 MD シミュレーションを並列計算機 CM-5 上に実装した結果を示し，計算負荷の平均化性能について議論する．5 章はまとめである．

2. 並列分子動力学法シミュレーション

本論文で扱う MD シミュレーションは古典 MD と呼ばれ，各分子に働く力を分子間相互作用から計算してニュートンの運動方程式に代入し，数値積分を行うことにより時間刻み（タイムステップ）ごとの各分子の位置と速度を計算する．分子間相互作用は多くの場合，分子間距離に依存するポテンシャル関数で定義する．本論文では，このポテンシャル関数に Lennard-Jones 型ポテンシャル

$$V(r) = 4\epsilon\left\{\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right\} \qquad (1)$$

を使用する．式中の r は分子間距離である．ϵ はポテンシャルの極小値の大きさ，σ は分子の直径を表し，物質によって異なる定数である．Lennard-Jones 型ポテンシャルでは分子間距離が増大すると相互作用の大きさが急速に零に近づくため，「力の切断距離」r_c を導入し，r_c 以上離れた分子間では相互作用を零として力の計算時間を短縮する．この操作を「力の切断」と呼び，MD では広く用いられている．

一般的には，MD はすべての分子の組合せについて相互作用を計算する．そのため，計算複雑度は N 個の分子に対して ${}_NC_2 = N(N-1)/2$ により，$O(N^2)$ である．力の切断は相互作用の計算を削減するが，単純な実装では $N(N-1)/2$ 組の分子間距離を計算するため，計算複雑度は $O(N^2)$ のままである．

次に，ドメイン分割法の概要を示す．**図 1** は，ドメイン分割法を 2 次元空間の MD シミュレーションに適用した例である．図 1 では正方形のセルを使用して，セルの大きさを力の切断距離 r_c よりも大きく定める．そして 1 次元トーラス結合の 4 PE を使用し，1 行のセルをドメインとして PE に割り当てる．セルの大きさが r_c よりも大きいため，分子間相互作用を計算するべき分子の組合せは，各セル自身と 8 近傍セル中に含まれる．そのため，図 1 では各 PE は 2 近傍の PE とだけデータを交換すればよい．3 次元空間のシミュ

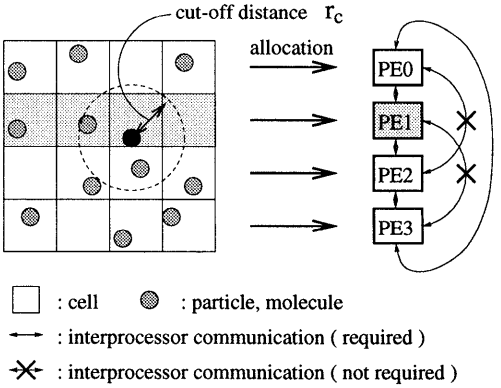


図 1 ドメイン分割法
Fig. 1 Domain decomposition method.

レーションの場合はセルは立方体とし，セルの大きさを力の切断距離よりも大きく定める．このとき，各セル自身と 26 近傍セル中の分子を参照すれば，相互作用の計算に必要な分子の組合せが得られる．このドメイン分割法によって計算量と通信量を削減することができる．

3 次元 MD シミュレーションにおけるドメイン分割法の計算複雑度を評価する．(分子間距離を計算する分子の組合せ数) $\geq$ (相互作用を計算する分子の組合せ数) であることから，分子間距離を計算する組合せ数によって計算複雑度を評価する．セルの 1 辺は無次元化した力の切断距離 r_c^* ($= r_c/\sigma$) に等しいものとし，分子はシミュレーション空間中に均等に分布すると仮定する．シミュレーション空間中の全セル数を C とすると，シミュレーション空間の体積は Cr_c^{*3} である．全分子数を N とすると，密度 ρ^* は

$$\rho^* = \frac{N}{Cr_c^{*3}}, \qquad (2)$$

で与えられる．各セル中には同数の分子が存在すると仮定し，1 つのセル中に存在する分子の数 N_{cell} は，

$$N_{cell} = r_c^{*3}\rho^*, \qquad (3)$$

で定義する．議論を簡単にするため，N_{cell} は整数であるものとする．このとき，C 個のセル中に N 個の分子が存在し，1 個のセル中に N_{cell} 個の分子が存在するため，

$$N = N_{cell}C = r_c^{*3}\rho^* C, \qquad (4)$$

の関係がある．1 個のセルに着目すると，分子間距離を計算する分子の組合せには，そのセル中での分子の組合せと近傍セル中の分子との組合せがある．全シミュレーション空間での分子間距離の組合せ数 I_{pairs} は

$$I_{pairs} = \left[\frac{N_{cell}(N_{cell}-1)}{2} + \frac{26N_{cell}^2}{2}\right]C, \qquad (5)$$

$$= \left[\frac{27}{2}N_{cell}^2 - \frac{1}{2}N_{cell}\right]C, \qquad (6)$$

$$= \left[\frac{27}{2}r_c^{*3}\rho^* - \frac{1}{2}\right]N, \qquad (7)$$

である．式 (5) の第 1 項は各セル内での N_{cell} 個の分子の組合せ数であり，第 2 項は各セル中の N_{cell} 個の分子と近傍の 26 セル中の N_{cell} 個の分子との組合せ数である．この I_{pairs} がドメイン分割法の計算複雑度を表す．式 (6) の C の係数では N_{cell} が上限を持つため，ドメイン分割法の計算複雑度は $O(C)$ である．式 (7) では，r_c^* は定数であり，ρ^* は上限値を持つ．そのため，ドメイン分割法の計算複雑度は $O(N)$ でもある．

次に，式 (7) の N の係数 I_{pairs}/N を評価し，I_{pairs}/N が N に対して十分小さいことを示す．Lennard-Jones 型ポテンシャルでは，r_c^* は 2.5 から 3.5 程度である．密度の最大値は面心立方格子の密度 $\rho^* \simeq 1.41$ 程度である．そのため，$r_c^* = 2.5$ で N_{cell} の最大値は 22 程度であり，$r_c^* = 3.5$ で N_{cell} の最大値は 60 程度である．I_{pairs}/N は $r_c^* = 2.5$ で 297 程度，$r_c^* = 3.5$ で 816 程度である．並列 MD シミュレーションでは $N \geq 10000$ の分子を扱うので，Lennard-Jones 型ポテンシャルの働く MD シミュレーションでは，I_{pairs}/N は N よりも十分小さい．

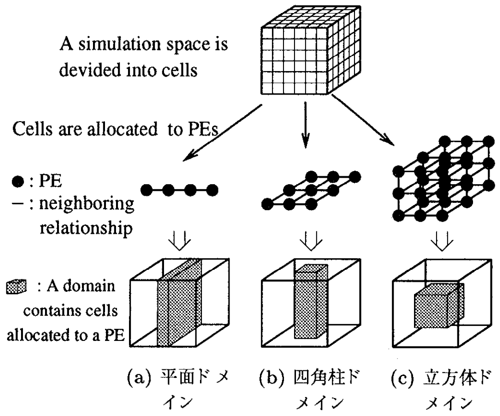


図 2 3 次元 MD シミュレーションの 3 種類のドメイン形状
Fig. 2 Three domain shapes in three-dimensional MD simulation.

3 次元空間の MD シミュレーションにドメイン分割法を適用する際，**図 2** に示すような 3 つのドメイン形状が可能である．我々は文献 3) および 4) でドメイン形状について議論し，64 PE 程度の並列計算機 CM-5 を使用して中規模シミュレーションを行う場合には，四角柱ドメインが最適であることを示した．そこで本論文では，図 2 (b) に示す四角柱ドメインを用いて DLB を議論する．四角柱ドメインでは，使用する PE 間で仮想的に 2 次元メッシュ結合網（仮想結合網）をつくり，シミュレーション空間を 2 次元で分割したドメインを PE に割り当てる．このとき，ドメインおよび PE は 8 近傍関係を持つ．

3. 動的負荷分散法

3.1 計算負荷の不均衡

MIMD 型並列計算機を用いた MD シミュレーションでは，1 タイムステップごとにすべての PE 間で同期をとる必要がある．しかし同期をとると，全 PE は最も処理時間の遅い PE が処理を終えるのを待つ必要がある．PE が持つ分子の数が不均等であるとき，分子を多く持つ PE は他の PE よりも 1 タイムステップ

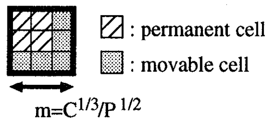


図 3 四角柱ドメインの長軸に垂直な断面図中の固定分配セルと可動セル
Fig. 3 The permanent cells and the movable cells in a cross-section of a square pillar domain.

の処理時間が大きくなり，同期によって全体の並列化効率は悪化する．

ここで，並列 MD シミュレーションで起こりうる計算負荷の不均衡の程度を簡単に評価する．2 章の議論より，ドメイン分割法を用いた MD シミュレーションの計算複雑度は $O(C)$ であった．式 (6) から得られる I_{pairs}/C の最大値と最小値が 1 セルあたりの計算負荷の範囲を表す．I_{pairs}/C の最小値は $N_{cell}=0$ のときに $I_{pairs}/C=0$ である．面心立方格子の分子密度 $\rho^*_{fcc}=1.41$ と $r^*_c=2.5$ を用いると，I_{pairs}/C の最大値は $I_{pairs}/C=6523$ である．すなわち 1 セルあたりの計算負荷は，0〜10^3 程度の範囲で変化する．

3.2 固定分配セルと動的負荷分散

DLB は，セルの再分配，すなわち近傍 PE 間でセルを移動することによって計算負荷移動を行う．しかし，セルの再分配に際しては，仮想結合網を保存する必要がある．ドメイン分割法では，ドメインの近傍関係と仮想結合網は等価である．セル移動によってドメインの近傍関係が変化すると，各 PE が通信を行う相手数の上限が全 PE 数になり，一般には PE 間通信のオーバヘッドは増大する．

仮想結合網の規則性を保存しながらセルの再分配を行うために，本論文では「固定分配セル」の概念を導入する．四角柱ドメインの長軸に垂直な断面図を用いて固定分配セルの定義を**図 3** に示す．四角柱ドメインの底面の大きさは，$m=C^{1/3}/P^{1/2}$ セルである．ドメインの断面中で 1 行分のセルおよび 1 列分のセルを固定分配セルとし，それ以外のセルを可動セルとする．可動セルは，セルの再分配によって，他の PE に移動可能である．一方，固定分配セルは他の PE に移動せず，各 PE は割り当てられたドメインの固定分配セルの計算をシミュレーションの最後まで担当する．

可動セルの移動について述べる．図 3 に示したドメインを 9 PE に割り当てた状態を**図 4** に示す．図 4 中の $PE(i,j)$ は PE のアドレスを示す．固定分配セルの形状から $PE(i,j)$ は，担当するドメインの可動セルを $PE(i-1,j-1)$，$PE(i-1,j)$，$PE(i,j-1)$ に送ることができる．また，$PE(i,j)$ は $PE(i,j+1)$，

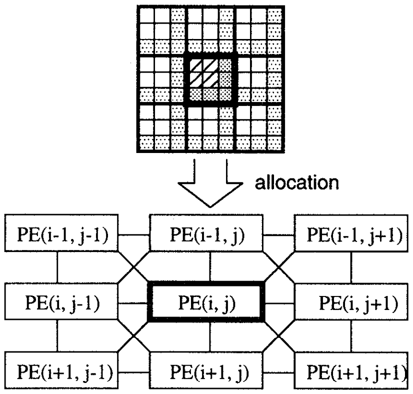


図 4 固定分配セルとプロセッサ要素へのセルの割当て
Fig. 4 An example of allocation cells to PEs.

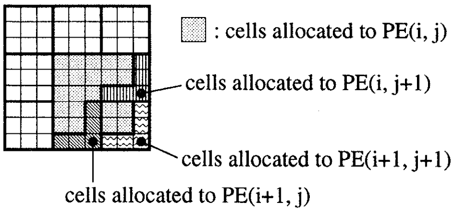


図 5 プロセッサ要素が担当可能な最大数のセル
Fig. 5 An result of cell redistribution $PE(i,j)$ has maximum number of cells.

$PE(i+1,j)$，$PE(i+1,j+1)$ から，各 PE が担当するドメインの可動セルを受け取ることができる．そして $PE(i,j)$ は，各 PE の計算負荷分布状況が変化したとき，以前に他の PE から受け取ったセルを元の PE に送り返すことができる．しかし，仮想結合網を保存するためには，上記以外のセル移動はできない．

仮想結合網を保存するセル移動によって，各 PE の通信量は増加する．通信量の増加が通信時間に与える影響について，付録 A.1 で議論する．

固定分配セルが仮想結合網の規則性を保存する簡単な例を**図 5** に示す．図 5 は，$PE(i,j)$ が $PE(i,j+1)$，$PE(i+1,j)$，$PE(i+1,j+1)$ からすべての可動セルを受け取った状態であり，$PE(i,j)$ は担当可能な最大数のセルを持っている．この状態でも PE の 8 近傍関係は保存されている．このように DLB では，仮想結合網を保存しながら計算負荷を平均化することができる．しかし，図 5 の状態であっても $PE(i,j)$ の計算負荷が 9 PE 中で最小であるときは，これ以上の計算負荷の平均化はできない．固定分配セルを用いた DLB では，可動セルの数とセル再分配の方向に制限があるために，計算負荷の平均化性能には限界がある．

3.3 動的負荷分散法の計算手順

ドメイン分割法に DLB を組み合わせた DLB ドメ

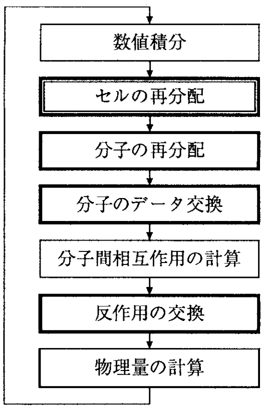


図 6 DLB-ドメイン分割法の計算手順
Fig. 6 Simplified structure of program with DLB-domain decomposition method.

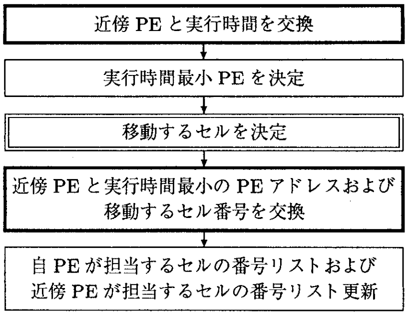


図 7 セル再分配の手順
Fig. 7 Procedures for cell redistribution.

```
1  FindSendCell(C_Movable, C_Received, PE_Fastest)
2  integer C_Movable(NOfMvble)
   : 担当するドメインの可動セル番号リスト
3  integer C_Received(NOfRcvd, PE_HlfNghbr)
   : 近傍 PE から受けとった可動セル番号リスト
4  if(PE_Fastest = {MyNode ∪ NE ∪ SW}) then
5      FindSendCell = NoCell
6  else
7      if(PE_Fastest = {W ∪ NW ∪ N}) then
8          FindSendCell = FirstLoc(C_Movable)
9      endif
10     if(PE_Fastest = {E ∪ SE ∪ S}) then
11         FindSendCell = FirstLoc (C_received
               , PE_Fastest)
12     endif
13 endif
14 return (FindSendCell)
```

図 8 移動するセルの決定法
Fig. 8 Algorithm to decide a moving cell.

イン分割法の計算手順を図 **6** に示す．図 6 の太枠部分は，近傍 PE 間のデータ交換を要する手順である．DLB ドメイン分割法の計算手順はセルの再分配を通常のドメイン分割法に加えたものであり，セルの再分配を除けば，DLB ドメイン分割法とドメイン分割法の計算手順は同様である．

DLB ドメイン分割法では，分子の新しい位置と速度を計算する数値積分の後に，セルの再分配を行う．ドメイン分割法において各 PE は，セル番号リストを用いてセルと対応付けながら，分子のデータを扱う．セルの再分配の段階ではこのセル番号リストのみを更新し，移動したセル中の分子のデータを移動するのは，次に行う分子の再分配の段階である．分子の再分配は，分子の位置から所属するべきセル番号を計算し，新しいセル番号に従って分子をセル間および PE 間で移動する．分子の再分配の際に，分子のセル間の移動とセルの再分配にともなう分子の移動を同様に扱うことができる．

セルの再分配の手順を図 **7** に示す．図 6 と同様に，近傍の PE 間でデータ交換を要する手順を太枠で示した．セルの再分配では，近傍 PE 中で前のタイムステップの実行時間が最小である PE を調べ，その PE に移動するセルを決定する．そして，実行時間最小の PE アドレスおよび移動するセル番号を近傍 PE と交換し，セル番号リストを更新する．セルの再分配は近傍 PE のみの実行時間データを使用して，近傍 PE 間で計算負荷移動を行う．そのため動的負荷分散を局所的に行うことができる．

移動するセルを決定するアルゴリズムを図 **8** に示す．

4〜5 行目 MyNode, NE, SW は各々自 PE，北東 PE，南西 PE を表す．図 4 中の表記に従うと，これらは $PE(i,j)$，$PE(i-1,j+1)$，$PE(i+1,j-1)$ である．これらの PE が $PE_{Fastest}$ である場合，セルの移動は行わない．

7〜8 行目 W, NW, N は各々西 PE，北西 PE，北 PE を表す．図 4 では，これらは $PE(i,j-1)$，$PE(i-1,j-1)$，$PE(i-1,j)$ である．これらの PE が $PE_{Fastest}$ である場合，$C_{Movable}$ の中から 1 セルを選択する．

10〜11 行目 E, SE, S は各々東 PE，南東 PE，南 PE を表す．図 4 では，これらは $PE(i,j+1)$，$PE(i+1,j)$，$PE(i+1,j+1)$ である．これらの PE が $PE_{Fastest}$ である場合，$C_{Received}$ の中で $PE_{Fastest}$ からかつて受け取ったセルの中から 1 セルを選択する．

図 8 において戻り値 *NoCell* はセルの移動を行わないことを表し，関数 *FirstLoc* は該当するセルがない場合に *NoCell* を返す．また，関数 *FirstLoc* は第 2 引数（PE のアドレス）があるときは，第 2 引数の PE

から受け取ったセル中でセルを選択する．

図8のセル移動の決定法によると，各PEはそれぞれの近傍PEにセルを移動するため，あるPEが1度に3つの近傍PEからセルを受け取ることがある．そのため，大きな計算負荷の違いが生じても速やかに平均化できる．もし計算負荷を移動しすぎた場合でも，次のタイムステップ以降において元のPEに送り返せば，計算負荷量の範囲を一定に保つことができる．

4. 動的負荷分散法の性能評価

4.1 並列計算機環境

DLBの計算負荷平均化の効果を検討するため，本論文では64PEの並列計算機CM-5[9]上に，並列MDシミュレーションを実装した．CM-5はPEに32MHzのSPARC microprocessorsと4つのベクトルユニットを使用した分散メモリ型並列計算機である．ベクトルユニットの性能は1基あたり16MHz, 32MFLOPSである．また各PEは，FPU（32MHz, 22MIPS, 4.2MFLOPS）および32MByteのメモリを持つ．結合網はFat tree型であり，1PEごとに2つのネットワークインタフェース（20MByte/second）を持つ．プログラムには，FORTRANの拡張言語であるCM FortranにメッセージB通信ライブラリCMMDを併用し，MIMD型処理を行った．本論文で用いる実行時間は，CMMDのCMMD_node_timer_busy関数を使用して計測した．

4.2 並列MDシミュレーションの物理条件

本論文で扱うMDシミュレーションでは，分子間相互作用にLennard-Jones型ポテンシャルを使用する．Lennard-Jones型ポテンシャルのパラメータには，アルゴンを模擬する数値を用い，力の切断距離 r_c^* は2.5を使用する．物理系は分子の数 N，シミュレーション空間の体積 V，および全エネルギー E を保存する．全エネルギーは運動エネルギーと位置エネルギーの和であるが，位置エネルギーが変化するために運動エネルギーが変化し，温度が一定にならない．そのため温度は50タイムステップごとに T_{ref}^* に補正する．数値積分にはベルレの速度形式[1]を使用した．1タイムステップの時間刻み Δt は0.064で，境界は周期境界条件とした．初期状態では，分子の位置は規則的である．

分子が時間とともに集中する状態を作り出すため，無次元化温度 $T_{ref}^* = 0.722$ で無次元化数密度 $\rho^* = 0.256$ の物理状態を用いる．無次元化温度の定義[1]によると，$T_{ref}^* < 1.0$ のとき，大部分の分子の運動エネルギーはポテンシャルの極小値の深さよりも小さい．このとき大部分の分子は，他の分子のポテンシャルの引力部分の極小点に位置すると，そのまま引力部分にとどまる．一方，無次元化数密度 $\rho^* = 0.256$ は，最大の無次元化密度（面心立方格子の密度 $\rho_{fcc}^* = 1.41$ 程度）よりも小さい．そのためここで用いた物理条件では，分子を規則的に配置した初期状態から離れると，分子の集中が起こる．

4.3 実験結果

DLBの効果を調べるため，従来のドメイン分割法を用いたシミュレーションとDLBドメイン分割法によるシミュレーションを実装し比較する．実験結果の典型例を示すため，2つのシミュレーション条件を用いる．

シミュレーション1 16PEを用いて8000分子を含む1728セルを計算した．1個のドメインは3×3×12セルを持つ．可動セルの割合は，ドメインの4/9である．

シミュレーション2 64PEを用いて17576分子を含む4096セルを計算する．1個のドメインは2×2×16セルを持つ．可動セルの割合は，ドメインの1/4である．

図9に，シミュレーション1における従来のドメイン分割法（DDM）とDLBドメイン分割法（DLB-DDM）の並列処理時間とタイムステップの関係を示す．縦軸は1タイムステップの並列処理時間であり，横軸はタイムステップである．DDMでは1000タイムステップ以降で並列処理時間が急速に増加するのに対して，DLB-DDMの並列処理時間はほぼ一定である．固定分配セルを用いたDLBは，シミュレーション1では計算負荷を平均化することができる．

図9のDDMとDLB-DDMのそれぞれについて，

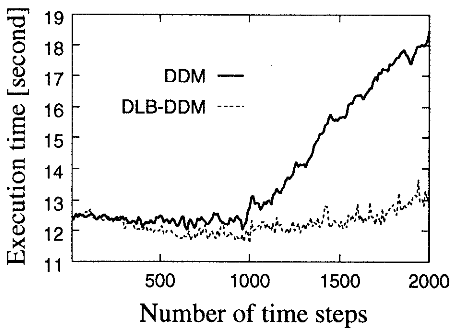
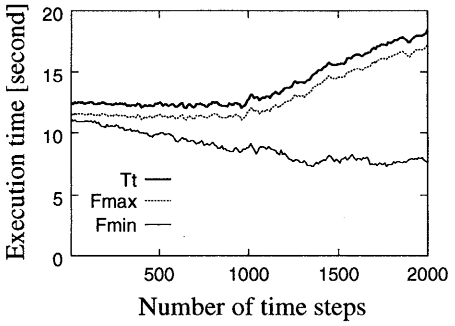
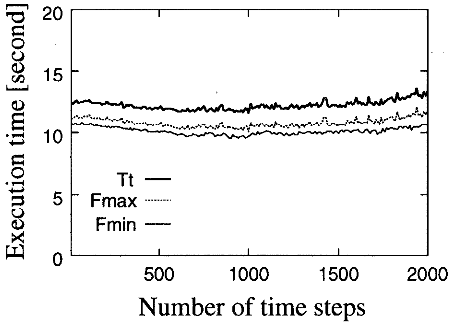


図9 ドメイン分割法とDLBドメイン分割法の並列実行時間とタイムステップの関係（16PE, $N = 8000$, $C = 1728$, $T_{ref}^* = 0.722$, $\rho^* = 0.256$）

Fig. 9 Execution time as a function of time step of domain decomposition method and DLB domain decomposition method on 16 PEs of CM-5 ($N = 8000$, $C = 1728$, $T_{ref}^* = 0.722$, $\rho^* = 0.256$).



(a) ドメイン分割法の実行時間



(b) DLB ドメイン分割法の実行時間

図 10 1タイムステップの実行時間と相互作用の計算時間（16 PE, $N = 8000$, $C = 1762$, $T^*_{ref} = 0.722$, $\rho^* = 0.256$）

Fig. 10 Execution time and force calculation times of domain decompsition method and DLB domain decomposition method on 16 PEs of CM-5 ($N = 8000$, $C = 1728$, $T^*_{ref} = 0.722$, $\rho^* = 0.256$).

1タイムステップの並列処理時間の内訳を**図 10** に示す．図中の量は次のとおりである．

Tt 1タイムステップの並列処理時間

Fmax 16 PE における最大の相互作用計算時間

Fmin 16 PE における最小の相互作用計算時間

各タイムステップで同期をとることから，DDM と DLB-DDM のどちらでも Tt は Fmax に依存して変化する．DDM の Fmax と Fmin の差は，タイムステップが進むに従って増大する．これは，多くの分子が少数の PE に集中して計算負荷が不均一になったことを示す．それに対して DLB-DDM では，Fmax と Fmin の差は，シミュレーション全体を通じて小さい．これは，DLB が計算負荷を平均化していることを示す．

図 11 はシミュレーション 2 における DDM と DLB-DDM の並列処理時間とタイムステップの関係を示す．シミュレーション 2 では可動セルの全セル数に占める割合が小さいので，DLB-DDM の計算負荷平均化性能はシミュレーション 1 よりも小さい．図 11 によると，600 タイムステップまでは DLB のオーバヘッドの効果が計算負荷平均化の効果よりも大きいために，DLB-DDM は DDM よりも実行時間が大きい．そして，DLB-DDM は 600 タイムステップから 1600 タイムステップまでは，実行時間を DDM よりも減らすことができる．しかし，1200 タイムステップ以降では，DLB-DDM と DDM の実行時間の差は小さくなりはじめる．これは，シミュレーション 2 において，計算負荷の不均衡が DLB の計算負荷平均化性能の限界を超えたことを示す．

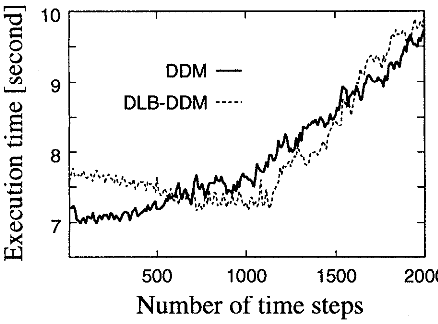


図 11 ドメイン分割法と DLB ドメイン分割法の並列実行時間とタイムステップの関係（64 PE, $N = 17576$, $C = 4096$, $T^*_{ref} = 0.722$, $\rho^* = 0.256$）

Fig. 11 Execution time as a function of time step of domain decompsition method and DLB domain decomposition method on 64 PEs of CM-5 ($N = 17576$, $C = 4096$, $T^*_{ref} = 0.722$, $\rho^* = 0.256$).

4.4 計算負荷平均化性能の定式化

固定分配セルを用いた DLB は，計算負荷平均化性能に限界があった．そこで本節では，DLB ドメイン分割法が有効である計算負荷集中の大きさについて議論する．DLB ドメイン分割法では，各 PE が近傍で計算時間最小である PE にセルを送るという方法で計算負荷を移動する．そのため計算負荷平均化可能な状態では，PE が担当可能な最大のシミュレーション空間（最大担当空間）の計算負荷量が平均の計算負荷量以上であると考えられる．本節では，分子が最も少ない最大担当空間の計算負荷量を定式化して，計算負荷の集中度を評価する．

ドメイン分割法を用いた MD の計算複雑度が $O(N)$ かつ $O(C)$ であることを利用し，定式化の議論を簡略化するために以下の仮定を置く．

(1) セルの形状は立方体であり，セルの 1 辺の大きさは r_c^* に等しい．

(2) シミュレーション空間全体で，分子を含まないセルは C_0 個であり，残りの $C - C_0$ 個のセルは，すべて $N/(C - C_0)$ 個の分子を含む．

(3) どの方向へもセル再分配が可能である．

(4) 分子を含まないセル数の割合をドメインごとに求めて得る最大値と，最大担当空間での分子を含まないセル数の割合は等しい．

さらに，シミュレーション空間は立方体であるものとし，シミュレーション空間の1辺のセル数 $C^{1/3}$ および仮想結合網の1辺のPE数 $P^{1/2}$ は整数であるものとする．そして，C' を最大担当空間のセル数，C'_0 を最大担当空間中の分子を含まないセル数とする．定式化のために，以下のパラメータを使用する．

m： 図3で導入した，四角柱ドメインの底面の1辺の長さ．セル数 C とPE数 P および m の間には次式の関係がある．

$$m = \frac{C^{1/3}}{P^{1/2}} > 1. \tag{8}$$

m によって，可動セルと固定分配セルの割合，および最大担当空間のセル数 C' が一意に定まる．

n： 最大担当空間における分子を含まないセルの集中度．次式で定義する．

$$n = \frac{C'_0/C'}{C_0/C} \geq 1. \tag{9}$$

平均化可能な計算負荷集中の定式化を行う．P 個のPEが N 個の分子を含む C セルを計算するとき，PEが持つ平均分子数は次式で表せる．

$$\frac{N}{P}. \tag{10}$$

最大担当空間中のセル数 C' は，1つの四角柱ドメインおよび近傍3PEの全可動セル数の総和である．

$$C' = [m^2 + 3(m-1)^2]C^{1/3}. \tag{11}$$

式(11)に分子を含むセルの割合および分子を含むセル中の分子の数を乗ずれば，最大担当空間中の分子数が得られる．最大担当空間中の分子を含まないセルの割合は平均の n 倍であることから，最大担当空間中での分子を含むセルの割合は次式で表せる．

$$\frac{C' - C'_0}{C'} = 1 - n \cdot \frac{C_0}{C}. \tag{12}$$

仮定(2)より分子を含むセル中の分子の数は

$$\frac{N}{(C - C_0)}, \tag{13}$$

である．最大担当空間中の分子数が，1PEの持つ平均分子数以上であれば，負荷の平均化が可能である．この条件は式(10)，式(11)，式(12)および式(13)を用いて，次の式で表せる．

$$[m^2 + 3(m-1)^2]C^{1/3} \times \left(1 - \frac{nC_0}{C}\right)\frac{N}{C - C_0} \geq \frac{N}{P}. \tag{14}$$

両辺 $\times\, P(C - C_0)/N$ より，

$$[m^2 + 3(m-1)^2]C^{1/3}P\left(1 - \frac{nC_0}{C}\right) \geq C - C_0. \tag{15}$$

左辺の $(1 - nC_0/C)$ を展開し，右辺に C_0 の項をまとめて

$$[m^2 + 3(m-1)^2]C^{1/3}P - C \geq \{n[m^2 + 3(m-1)^2]\frac{P}{C^{2/3}} - 1\}C_0. \tag{16}$$

ここで両辺に m^2/C を乗じ，式(8)により $C^{1/3}/P^{1/2}$ を m に置きかえると以下のように変型できる．

$$[m^2 + 3(m-1)^2] - m^2 \geq \{n[m^2 + 3(m-1)^2] - m^2\}\frac{C_0}{C} \tag{17}$$

$$\Updownarrow$$

$$\frac{3(m-1)^2}{n[m^2 + 3(m-1)^2] - m^2} \geq \frac{C_0}{C}, \tag{18}$$

$$n[m^2 + 3(m-1)^2] - m^2 > 0. \tag{19}$$

式(18)の左辺を $f(m,n)$ とおいて，次式を得る．

$$f(m,n) = \frac{3(m-1)^2}{m^2(n-1) + 3n(m-1)^2} \geq \frac{C_0}{C}, \tag{20}$$

$$f(\infty,n) = \lim_{m\to\infty} f(m,n) = \frac{3}{4n-1}. \tag{21}$$

$f(m,n)$ は計算負荷平均化が可能な C_0/C の上界を表す．その理由は主に2つである．式(20)導出のための仮定(3)に反して，DLBではセルの移動方向は3方向のみである．さらに，式(20)の導出は最大担当空間全体のセル数のみを考慮した議論であり，DLBにおいてセルの移動方向ごとに移動可能なセル数の制限があることを考慮していない．これらの理由のため，DLBの計算負荷平均化性能は必ず $f(m,n)$ よりも小さい．

式(20)によると，$f(m,n)$ は四角柱ドメインの底面の大きさ m および分子を含まないセルの集中度 n に依存し，使用PE数 P に直接の依存性を持たない．式(21)はドメインの大きさが無限大であって固定分配セルのドメインに占める割合が零である場合を表す．すなわち $f(\infty,n)$ は，固定分配セルの影響が無視できる場合の上界であって，$f(\infty,n)$ を超える負荷集中を完全に平均化することは不可能である．

表1は，式(20)と式(21)にいくつかの m および n の数値を代入した結果である．m が増加するとドメインの底面が大きくなってドメイン中での固定分配セルの割合が小さくなり，DLBの計算負荷平均化性

表 1 計算負荷の平均化が可能な C_0/C の最大値の上界
Table 1 Upper bound of C_0/C for making even computational load.

$f(m,n)$	$n=2$	$n=3$	$n=4$
$m=2$	0.30	0.18	0.13
$m=3$	0.36	0.22	0.16
$m=4$	0.39	0.24	0.17
$m\to\infty$	0.43	0.27	0.20

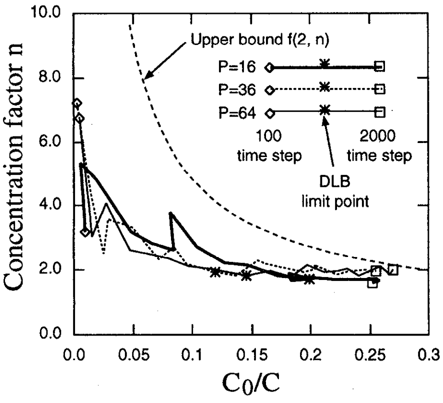
能が大きくなる．表1によると，n を固定して m を増加させると $f(m,n)$ が大きくなり，計算負荷平均化性能が向上することを示す．このことから，式(20)によっても，ドメインを大きくすると計算負荷平均化性能が向上することが分かる．また，m が増加すると $f(m,n)$ は急速に $f(\infty,n)$ に漸近し，$f(4,n)$ は $f(\infty,n)$ の約90%である．つまり四角柱ドメインの底面の1辺が4セルであれば，固定分配セルの影響が無視できる場合での計算負荷平均化性能の90%を実現可能である．

n の増加は最大担当空間における分子を含まないセルの集中度が大きくなることを表し，定式化の仮定(4)により，特定のPEへの計算負荷の集中度が増加することを示す．そのため n が大きくなると，DLBの計算負荷平均化性能が小さくなると考えられる．実際に表1において，n が大きくなると同じ m に対して $f(m,n)$ が小さくなり，計算負荷平均化性能の低下が起こる．

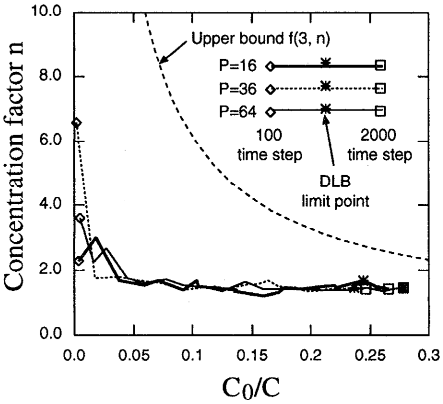
4.5 並列MDシミュレーションの計算負荷集中

定式化で得られたDLBの理論的な計算負荷平均化性能と，並列MDシミュレーションの実行から得られたDLBの計算負荷平均化性能を比較する．分子の入らないセルの集中度 n と分子の入らないセルの割合 C_0/C の関係を表す n–C_0/C 空間を定義する．実行した並列MDシミュレーションにおいて，100タイムステップごとに n および C_0/C を計算して n–C_0/C 空間にプロットすると，並列MDシミュレーションにおける計算負荷集中度の軌跡が得られる．

図12(a)，(b)は，$m=2$ および $m=3$ の並列MDシミュレーションと，$f(2,n)$ および $f(3,n)$ の n–C_0/C 空間中の軌跡である．並列MDシミュレーションの結果は，PE数 $P=16$，$P=36$，$P=64$ を用いた3つの軌跡を示す．タイムステップの進行に従って，シミュレーションの軌跡は n–C_0/C 空間中で大まかに左から右に移動する．図12中のDLB limit pointは，DLBドメイン分割法において，一定または減少傾向にあった並列実行時間が増加を始めるタイムステップでの n–C_0/C 空間中の点である．DLB



(a) $m=2$



(b) $m=3$

図 12 計算負荷の平均化可能な C_0/C の上界と並列MDシミュレーションの軌跡
Fig. 12 Upper bound of C_0/C and trajectories of parallel MD simulation.

limit pointよりも左側の軌跡は，実装において計算負荷の平均化が可能であった領域を示す．物理条件が同一であるため，P が異なる場合でも，並列MDシミュレーションの軌跡は非常に類似している．$m=2$ および $m=3$ のどちらの場合も，実装において計算負荷の平均化が可能であった領域は，上界 $f(2,n)$ および $f(3,n)$ を満たす．

5. まとめ

3次元空間のMDシミュレーションは2次元空間に比べ，PEの近傍関係の保存および移動する計算負荷の粒度などの問題があるために，並列計算における計算負荷の平均化が困難であった．本論文ではPEの近傍関係を一様に保つため，固定分配セルの概念を動的負荷分散法に導入し，ドメイン分割法のセルを計算負荷の単位として動的負荷分散を行った．動的負荷分散

法を組み合わせたDLBドメイン分割法を並列計算機CM-5上に実装した結果，DLBドメイン分割法は高い計算負荷平均化効果を得ることができた．しかし，ドメイン中に固定分配セルが占める割合が大きい場合には，分子が非常に集中すると計算負荷平均化の効果が低下した．そこで平均化可能な計算負荷集中度の上界を定式化し，動的負荷分散法の計算負荷平均化性能を議論した．

さらに計算負荷の平均化可能な範囲を詳細に議論するためには，分子の密度を変化させたシミュレーションを複数行い，計算負荷の平均化可能な範囲を詳細に調べる必要がある．

謝辞 本研究の一部は文部省科学研究費補助金によって行われた．

参考文献

1) Heermann, D.W.: シミュレーション物理学，シュプリンガー・フェアラーク東京 (1990).
2) Beazley, D.M., Lomdahl, P.S., Grønbech-Jensen, N., Giles, R. and Tamayo, P.: Parallel Algorithms for Short-range Molecular Dynamics, *Annual Reviews in Computational Physics*, Vol.3, pp.119–175, World Scientific (1995).
3) Hayashi, R. and Horiguchi, S.: Domain Decomposition Scheme for Parallel Molecular Dynamics Simulation, *Proc. HPC Asia '97*, pp.595–600 (1997).
4) 林 亮子：巨大分子の超並列分子動力学法シミュレーションに関する研究，博士論文，北陸先端科学技術大学院大学 (1998).
5) Brugé, F. and Fornili, S.L.: Concurrent molecular dynamics simulation of spinodal phase transition on transputer arrays, *Computer Physics Communications*, Vol.60, pp.31–38 (1990).
6) Kohring, G.A.: Dynamic load balancing for parallelized particle simulations on MIMD computers, *Parallel Computing*, Vol.21, pp.683–693 (1995).
7) 林 亮子，堀口 進：並列分子動力学シミュレーションにおける動的負荷分散法，並列処理シンポジウム JSPP '96 論文集，Vol.96, pp.81–88 (1996).
8) Hayashi, R. and Horiguchi, S.: Parallelized Simulation of Molecular Dynamics by Domain Decomposition Strategy, *Proc. 1st World Congress on Systems Simulation*, pp.353–358 (1997).
9) Frontier 利用者の会（編）：北陸先端科学技術大学院大学情報科学センター利用の手引第1版 (1994).

付 録

A.1 DLBドメイン分割法における通信量増加

DLBドメイン分割法においてはセルの移動の結果，PE間の通信量が増加する．本付録では，DLBドメイン分割法の通信量の増加量を導出し，DLBドメイン分割法の通信量増加が通信時間に及ぼす影響を評価する．

任意の2PE間の同期通信時間を

$$T_s + (\text{境界セル数}) \times W_{cell} \times T_t, \tag{22}$$

で表せるものと仮定する．ここで T_s は同期通信のスタートアップ時間，W_{cell} は1個のセルに付随するデータの平均語数，T_t は1語あたりのデータ転送速度である．ここでは4バイトを1語とする．多くのメッセージ通信型並列計算機においては $T_s > T_t$ であり，通信時間への影響は T_s の方が T_t よりも大きい．本論文で使用している並列計算機環境では $T_s \simeq 300T_t$ である．1個のセルに付随する平均のデータ量 W_{cell} は，セル中に存在する分子の数 N_{cell} を用いた次式

$$W_{cell} = (\text{次元数}) \times N_{cell}, \tag{23}$$

によって評価する．本論文の物理条件では $N_{cell} < 5$，次元数は3であるため，$W_{cell} = 15$ 語とする．

四角柱ドメインを用いたドメイン分割法の通信量を導出する．図4の状態では，各ドメインの底面の大きさは m セル，各ドメインの長軸の大きさは $C^{1/3}$ セルである．そして $PE(i,j)$ は周囲のPEと**表2**のような通信を行う．これらの総和から，ドメイン分割法の通信セル数 C_{DDM} は次式で表せる．

$$C_{DDM} = 4(m+1)C^{1/3}. \tag{24}$$

DLBドメイン分割法の境界セル数を，近傍ドメインとの境界に接する最大担当空間中のセル数によって評価する．DLBドメイン分割法の最大の通信セル数 C_{DLBmax} は，本文中の図5において $PE(i,j)$ の担当するセル中で，近傍ドメインとの境界に接するセル数の総和である．**表3**に図5における $PE(i,j)$ の境界セル数をまとめた．表3から，C_{DLBmax} は次式で表せる．

$$C_{DLBmax} = 4(3m-1)C^{1/3}. \tag{25}$$

以上により，通信量の増加分をセル数で表すと次式が得られる．

$$C_{DLBmax} - C_{DDM} = 8(m-1)C^{1/3}. \tag{26}$$

式(26)に1セルの平均データ量 W_{cell} を乗すれば，通信量の増加分の語数が得られる．そこで，本文中で

表 2 ドメイン分割法でプロセッサ要素が交換するデータ量
Table 2 Number of cells to be transferred on domain decomposition method.

通信相手 PE アドレス	境界セル数
$PE(i-1,j-1), PE(i-1,j+1)$, $PE(i+1,j-1), PE(i+1,j+1)$	$C^{1/3}$
$PE(i-1,j), PE(i,j-1)$, $PE(i,j+1), PE(i+1,j)$	$mC^{1/3}$

表 3 DLB ドメイン分割法で最大担当空間を持つプロセッサ要素が交換するデータ量
Table 3 Number of cells to be transferred with maximum size of domain on DLB domain decomposition method.

通信相手 PE アドレス	境界セル数
$PE(i-1,j-1)$	$C^{1/3}$
$PE(i-1,j), PE(i,j-1)$	$(m+1)C^{1/3}$
$PE(i-1,j+1), PE(i+1,j-1)$	$mC^{1/3}$
$PE(i,j+1), PE(i+1,j)$	$(3m-2)C^{1/3}$
$PE(i+1,j+1)$	$(2m-3)C^{1/3}$

使用したシミュレーションについて，通信量の増加分の語数を評価する．

$P=16$，$C=1728$，$W_{cell}=15$ の場合

$$(C_{DLBmax}-C_{DDM})W_{cell}=2880. \qquad (27)$$

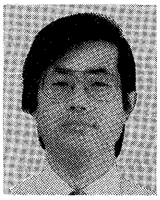
$P=64$，$C=4096$，$W_{cell}=15$ の場合

$$(C_{DLBmax}-C_{DDM})W_{cell}=1920. \qquad (28)$$

ドメイン分割法と DLB ドメイン分割法では，通信相手 PE 数は同じであるため，式 (27) および式 (28) に T_t を乗ずると，DLB で増加する通信時間が得られる．$T_s \simeq 300T_t$ であるため，本論文のシミュレーションでは，DLB ドメイン分割法における通信時間増加は T_s の 10 倍以内であることが分かる．一方，セル移動においてドメインの近傍関係を保存しないセル移動を許す場合は，通信相手数の上限は全 PE 数 P である．そのため，多数の PE を使用する MD シミュレーションでは，DLB ドメイン分割法が有効である．

(平成 10 年 8 月 31 日受付)
(平成 11 年 3 月 5 日採録)

林　亮子（正会員）
1970 年生．1993 年東北大学理学部物理学科卒業．1998 年北陸先端科学技術大学院大学情報科学研究科博士後期課程修了．1998 年 4 月より北陸先端科学技術大学院大学情報科学研究科助手．物理シミュレーションの並列化手法の研究に従事．IEEE 会員．



堀口　進（正会員）
昭和 27 年生．昭和 51 年東北大学工学部通信工学科卒業．昭和 56 年東北大学大学院工学研究科博士課程修了．昭和 56 年東北大学工学部助手，昭和 61 年 6 月～62 年 7 月米国 IBM ワトソン研究所・客員研究員として並列計算アルゴリズムの研究に従事．昭和 64 年東北大学工学部助教授．平成 3 年北陸先端科学技術大学院大学情報科学研究科教授，現在に至る．この間，並列処理，超並列システム，ウェーハ規模集積システム，並列アルゴリズム，マルチメディア統合システムに関する研究を行う．昭和 64～平成 7 年 IEEE 学会 WSI 国際会議組織委員会委員．平成 4～7 年同国際会議アジア地域議長．IEEE シニア会員，電子情報通信学会会員．