

5D-1

超並列計算機用言語 Bee-Fortran の
通信ライブラリの最適化方式の検討

杉森 英夫[†] 小前 晋[†] 大谷 浩司[‡] 渦原 茂[§] 安村 通晃[§]
[†]住友金属工業(株) [‡](有)AXE [§]慶應義塾大学

1 はじめに

現在我々は超並列計算機用言語 Bee-Fortran[2]を開発中である。Bee-FortranはHPF(High Performance Fortran)[3]に準拠したものである。Bee-Fortranの配列代入文などではPE間でのデータ通信が必要である。PE間通信のオーバーヘッドは実行性能に大きく関係する。そこで、PE間の通信量を減らすために、コンパイル時の静的な解析によりデータの配置の最適化は可能であるが、静的に最適化できない実行時のPE間通信に関しては、データ通信ライブラリを最適化することにより通信のオーバーヘッドを小さくできる。

本稿では、実行時のデータのPE間通信のオーバーヘッドを小さくするための最適化について、配列データの配置方式、通信パターンに応じた最適な通信方式について述べる。

2 Bee-Fortranの配列データ

Bee-Fortranは超並列計算機SM-1[1]をターゲットにした言語である。SM-1は2次元メッシュ(NEWS), shuffle-exchangeのグローバルネットワークなどを備えた32×32のPEで構成される分散メモリ型超並列計算機である。Bee-Fortranの配列データはユーザの指定がない限り自動的にPEへ配置される。ユーザは、配列要素をPEへ一つずつ番号順に割り当てていくサイクリック分割、配列要素をPE数で割った数ずつPEへ番号順に割り当てていくブロック分割、配列要素を指定された数ずつPEへ割り当てていく、ブロックサイクリック分割、の配置方法を指定することができる。

3 配列データの通信

Bee-Fortranでは実行時に配列代入文で左辺と右辺の対応している配列要素が同じPE上に無い場合など、配列のコピーが必要となる。配列データのコピーが必

The communication optimization scheme of Bee-Fortran libraries for a massively parallel machine

Hideo SUGIMORI[†], Susumu KOMAE[†], Koji OTANI[‡], Shigeru UZUHARA[§], Michiaki YASUMURA[§]

[†]SUMITOMO Metal Industries, LTD.

[‡]AXE Inc.

[§]KEIO University

要な時、コンパイラは配列のコピーを行なうライブラリを呼び出すオブジェクトを出力する。ライブラリでは、通信先、通信元のアドレス、通信先PE番号、通信サイズなどを通信ルーチンに渡し、shuffle-exchangeのグローバルネットワークを用いて配列データを通信することによりコピーを行なう。グローバルネットワークを用いた通信では通信を行なう際のデータの衝突、通信ルーチン呼び出しなどのオーバーヘッドによって通信コストは大きくなる。そこで、ライブラリ内で配列データのアクセスパターンを動的に解析し、最適な通信を行なうことにより通信コストを小さくする方法として以下のような通信の最適化を検討した。

- NEWSのネットワークを用いた最適化
- データのブロック化による最適化

3.1 NEWS通信を用いた最適化方式

配列データのアクセスパターンがNEWSのネットワークに沿っている場合、NEWSの通信路を通して隣接しているPEからデータを同じ方向に同時に通信するNEWS通信を用いることができる。NEWS通信を用いることによりグローバルな通信を行なう時のデータの衝突を避け通信コストを小さくすることができる。NEWS通信では通信先PE番号の代わりに通信方向と通信距離などを与えればよい。NEWS通信が使える場合は、対応している配列の軸の配置によって以下のような場合が挙げられる。

- コピー先、コピー元の対応する軸が同じ配置の場合(ブロック、サイクリック)

```
real a(100), b(100)
a(1:50:2) = b(51:100:2)
```

これらのNEWS通信を用いることができる場合では、コピー先、コピー元配列の対応した軸に対するtriplet添字の指定に従ったアクセスパターンによって通信元各PEに対する通信先PEの相対的な位置が全て同じである。

また対応している軸の配置が異なる場合であってもNEWS通信を用いることが可能な場合もある。

- コピー先、コピー元の対応する軸が異なる配置(ブロックとサイクリック)の場合、

```
real a(100), b(100)
distribution a(cyclic), b(block(3))
a(1:33) = b(1:100:3)
```

上の例では、コピー先、コピー元の対応する軸が異なる配置であるが、配列添字のアクセスパターンによって、NEWS通信が可能となっている。

2次元配列の場合でもそれぞれの対応している軸に関してアクセスパターン、配置が同じであればNEWS通信を用いることが可能である。

3.2 データのブロック化による最適化方式

次にコピーされる配列要素が連続した領域にある場合、その領域を一つのブロックとみなして通信を行なう。通信ルーチンにそのブロックの大きさをアドレスを与えることにより呼びだしなどのオーバーヘッドを減らすことが可能になる。

データをブロック化して通信を行なうには、

- 連続した領域を一つのブロックとして通信を行なう。
 - 同一 PE 上に配置された軸に対するアクセスを一つのブロックとする。
 - 同一 PE 上にブロック配置された連続した配列要素を一つのブロックとする。
- テンポラリブロックにバッファリングして、これを一つのブロックとして通信を行なう。

といった方法が考えられる。

1. 同一 PE 上に配置された軸に対するアクセスを一つのブロックとして通信を行なう場合

Bee-Fortran では特に指定が無い場合 3 次以上の配列要素は図 1 の様に同じ PE に配置される。ここで、3 次以上の軸に関してその添字が部分的に指定されてなければそれらの軸要素を一つのブロックとみなして通信を行なうことができる。

```
real a(100,100,100), b(100,100,100)
b(2,2,:) = a(1,1,:)
```

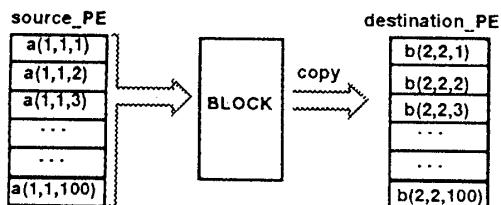


図 1: 同一 PE 上に配置された軸に対する通信

2. 同一 PE 上にブロック配置された連続した配列要素を一つのブロックとして通信を行なう場合

配列がブロックで配置されている場合は、連続した配列要素が同一 PE 上に配置されているため、PE 上の連続した要素をそのまま一つのブロックとして通信を行なうことができる。

3. テンポラリブロックを用いる場合

コピーされる配列データが triplet 添字の指定によって連続した領域に無い場合でも、コピー先の領域が連続しているなら、コピーされるデータを各 PE ごとにテンポラリブロックにバッファリングすることにより、一つのブロックとして通信を行なうことができる。

```
real a(100,100), b(100,100)
b(1:4) = a(1:6:2)
```

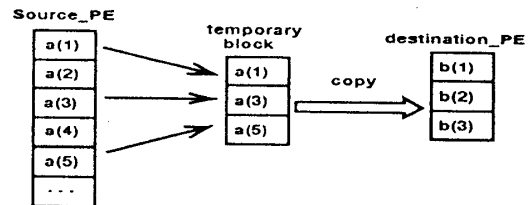


図 2: テンポラリブロックを用いた通信

4 おわりに

本稿では、Bee-Fortran の通信ライブラリの通信方式の最適化について述べてきた。Bee-Fortran での配列代入文での配列に対する triplet 添字を動的に解析することにより、最適な通信を行なうことで通信のオーバーヘッドをかなり軽減できる方式を明らかにした。

参考文献

- [1] 松田, 湯淺, SIMD 型超並列計算機 SM-1(仮称)の概要とその性能, 情報処理学会研究報告, 92-ARC-87, August, 1992
- [2] 大谷, 小前, 杉森, 渦原, 安村, 超並列計算機用 Fortran コンパイラの設計と試作, 電子情報通信学会研究報告, COMP92-94, May 11, 1993
- [3] High Performance Fortran Language Specification Version 1.0, High Performance Fortran Forum, May 3, 1991