

スーパーデータベースコンピュータ SDC における バケット平坦化機能を有するオメガネットワーク

5H-7

小川泰嗣(リコー 中央研究所)

喜連川優(東京大学 生産技術研究所)

1 はじめに

スーパーデータベースコンピュータ SDC は複数の CPU が密結合した処理モジュール (PM) を相互結合網により結合したハイブリッドアーキテクチャを採用した並列データベースマシンである [1]。SDC では、バケット分散並列結合演算法を採用する [2]。バケット分散方式では、各バケットを全 PM 上に均一に分配する必要がある。バケットを PM に均一分配する機能をバケット平坦化機能と呼ぶ。本発表では、SDC の相互結合網に用いるバケット平坦化機能を有するオメガネットワークについて説明する [3]。さらに、シミュレーションによる性能評価を報告する。

2 バケット分散並列結合法におけるバケット平坦分布

まず、SDC で採用されるバケット分散方式による並列ハッシュ結合法を簡単に説明する [2]。バケット分散方式では、バケットを PM にデータ分布に基づいて動的に割り当てる。そのため、PM 間で負荷を一定に調整でき、データ分布が不均一な場合にも効率の処理が実現される。その際、各バケットは複数の PM に分散格納される。各 PM に格納されるバケットの部分サブバケットと呼ぶ。

分散格納されていたバケットは、そのバケットの結合演算時に割り当てられた PM に収集される。多数の PM が無秩序にサブバケットを読出すとアクセス競合が発生するため、巡回的にバケット収集を行なう。しかし、サブバケットの大きさが不均一であれば、パイプラインの擾乱により処理速度が低下する。逆に、サブバケットの大きさが均一であれば、擾乱は発生せず効率の処理が実現できる。すなわち、バケット分散方式では、各バケットについてサブバケットの大きさが等しいことが要求される。このような分布をバケット平坦分布と呼ぶ。

3 バケット平坦化機能を有するオメガネットワーク

SDC では、バケット分散並列結合演算を実現するため、相互結合網に用いるオメガネットワークにバケット平坦化機能を埋め込む。

3.1 オメガネットワーク

オメガネットワークは多数の 2 入力 2 出力 (以下 2×2) スイッチング装置 (SU) から構成される。ある段の次の段はシャッフル交換によって結合される。ネットワークの大きさが $N \times N$ の場合、段数は $\log_2 N$ 、各段の SU 数は

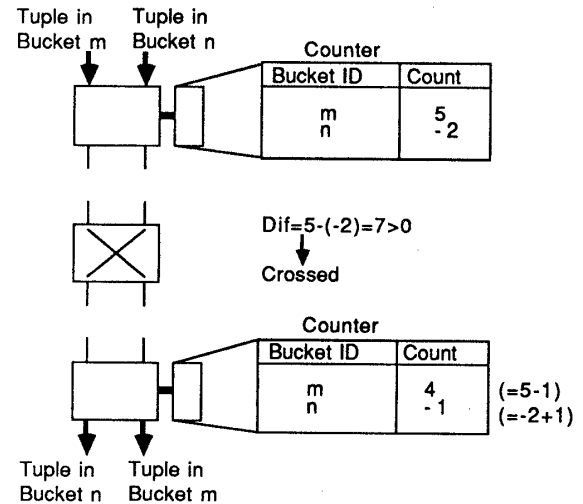


図 1: スイッチングユニットの動作例

$N/2$ である。各 SU は 2×2 のクロスバスイッチで、Straight と Crossed の 2 つの状態をとる。

3.2 分散制御によるバケット平坦化方式

バケット平坦化のためには、各タプルが属するバケットの PM 上の分布を調べ、そのバケットに関するサブバケットの大きさが最小の PM に転送すれば良い。我々が採用した分散制御では、PM がタプルの行き先を決定するのではなく、ネットワークの各 SU がバケットの分布の局所的な情報から自律的に状態を決定する。分散制御には、PM 数が多い場合にも適用可能、実装が簡単等の利点がある。以下で、SU の状態決定法を説明する。

各 SU に左右 2 つのポートから出力されたタプル数の差をバケットごとに記録するカウンタを用意する。全カウンタは 0 に初期化され、以後左のポートから出力されたタプルの属するバケット用のカウンタはインクリメント、右のポートから出力されたタプルの属するバケット用のカウンタはデクリメントされる。カウンタの値を $D(X)$ 、 X_{left} と X_{right} を左右の入力ポートに到着したタプルの属するバケットとすれば、 $Dif = D(X_{left}) - D(X_{right})$ は SU に入力された 2 つのタプルが属するバケットの出力の左右のポートによる偏りを表す。従って、SU の状態を Dif が正ならば Crossed、負ならば Straight とすればよい。図 1 に SU の動作例を示す。

An Omega Network Embedding the Bucket Flat Distribution Mechanism for the Super Database Computer (SDC)
Yasushi Ogawa (R & D Center, RICOH) and Masaru Kitsuregawa (Institute of Industrial Science, University of Tokyo)

4 シミュレーションによる性能評価

4.1 シミュレーションモデル

バケット分布の平坦度を次に定義する平均標準偏差で

評価する。 D_{ij} をバケット i の PM j に格納されているサブバケットの大きさとする、平均標準偏差は次式により計算される。

$$\bar{\sigma} = E\left\{\frac{1}{B} \sum_{i=1}^B \sqrt{\frac{1}{N} \sum_{j=1}^N D_{ij}^2 - \left(\frac{1}{N} \sum_{j=1}^N D_{ij}\right)^2}\right\}$$

$E\{\}$ は多数の試行に対する平均値を表す。 $\bar{\sigma} = 0$ がバケット平坦分布に相当、大きいほど分布は不均一となる。

タブルの属するバケットは次の確率分布の一方に従って決定される。第1の均一分布では、 $PM_i (i = 1, \dots, N)$ に格納されていたタブルがバケット $b (b = 1, \dots, B)$ である確率 $p_i(b)$ は $p_i(b) = 1/B$ である。第2のたんざく分布では、確率 $p_i(b)$ は次式で与えられる。

$$p_i(b) = \begin{cases} \frac{N}{B} & : (i-1) \times \frac{B}{N} < b \leq i \times \frac{B}{N} \\ 0 & : \text{otherwise} \end{cases}$$

各シミュレーションでバケット平坦化機能の効果を見るため、上に示した確率分布の平均標準偏差を求めた（初期標準偏差と呼ぶ）。

4.2 シミュレーション結果

シミュレーションでは、処理モジュール数 N 、PM ごとのタブル数 T およびバケット数 B の3つのパラメータを変化させ、性能評価を行なった。

処理モジュール数を変化させた場合: $T = 8k, B = 128$ に固定し、処理モジュール数を変化させた場合の結果を図2に示す。ネットワークが大きくなるほど初期標準偏差は大きくなり、 $N = 64$ で平均標準偏差は10（均一分布）、600（たんざく分布）程度となる。これに対し、本オメガネットワークによりどちらの分布に対しても、 $N = 64$ で平均標準偏差は0.6程度と非常に小さな値となる。

PM ごとのタブル数を変化させた場合: $N = 8, B = 128$ に固定し、PM ごとのタブル数を変化させた場合の結果を図3に示す。先ほどと同様バケット平坦化機能が有効であり、本ネットワークによりタブル数に関係なく平均標準偏差をほぼ一定にできる。

バケット数を変化させた場合: 前の結果に示されたように平均標準偏差はPM ごとのタブル数に影響されるため、PM ごとのタブル数はバケット数に比例させる。 $N = 8, T = 64 \times B$ （バケット数の64倍）とした場合の結果を図4に示す。本ネットワークによりバケット数に関係なく平均標準偏差を小さな値にできる。

5 おわりに

本発表では、SDC のモジュール間結合網に採用するバケット平坦化機能を有するオメガネットワークについて報告した。分散制御でバケット分散並列結合演算法に要求されるバケット平坦化機能を実現した。シミュレーションによる性能評価の結果、その有効性が確認できた。

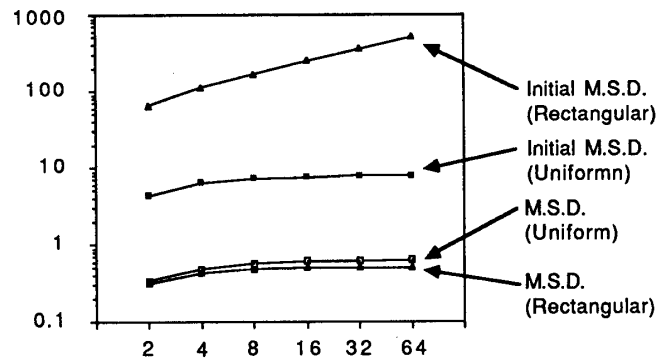


図 2: 処理モジュール数と平均標準偏差の関係

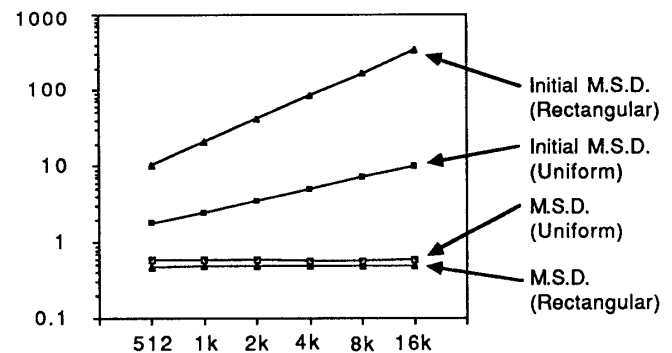


図 3: PM ごとのタブル数と平均標準偏差の関係

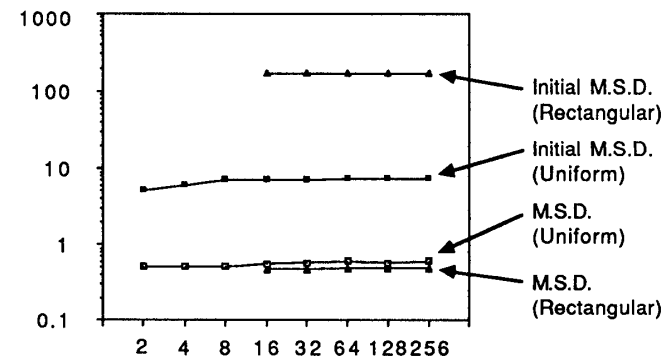


図 4: バケット数と平均標準偏差の関係

参考文献

- [1] 楊維康他. “スーパーデータベースコンピュータ SDC のアーキテクチャ”. 第 39 回情処全国大会.
- [2] 小川泰嗣, 喜連川優. “スーパーデータベースコンピュータにおけるバケット分散並列結合演算法とその性能予測”. 第 39 回情処全国大会.
- [3] 喜連川優, 小川泰嗣. “バケット平坦化機能を有するオメガネットワーク”. 情報処理学会論文誌, Vol. 30.