5H−7

スーパーデータベースコンピュータ(SDC)における バケット平坦化網の可変長データに対する制御方式

相場 雄一　平野 聡　喜連川 優　高木 幹雄
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持つ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でも効率的処理が実現される。それには各バケットを全PMに均一に分配する必要があるが、この際各バケットは複数のPMにサブバケットとして分散格納される。このサブバケットの大きさが等しいことが必要となり、これをバケット平坦化と呼ぶ。SDCではこのための機能を相互結合網に持たせたアルゴリズムを考え、固定長データについてはシミュレーションにより良好な結果を得ている[3, 4, 5, 6]。本発表では、データ長が可変の場合のスイッチングアルゴリズム及びシミュレーションによる性能評価を報告する。

2　バケット平坦化網の制御方式

結合網は多数の2入力2出力スイッチング装置(SU)から構成されたオメガ網である。SDCの平坦化網は、個々のSUがStraightとCrossedという2つの状態のいずれかを自律的に決定し、全体的に平坦化を実現しようというものである。固定長データに対する制御方式については文献[4, 5]にゆずるが、可変長データに対する制御方式も基本的な考え方は同様である。

まず、各SUにバケット毎のカウンタを用意し、このカウンタの状態によってSUの状態を決定する。カウンタは初期値0とし、以後タプルをSUのある一方の出力ポートに出力すればカウンタ値を増加、他方のポートに出力すれば減少するように決める。そして、SUの状態はこのカウンタ値を0に近付けるように決定される。このカウンタ値の増減の仕方によって次の2つのアルゴリズムが考えられる。

- タプル数による制御：1タプル出力ポートに送る毎に、カウンタ値を1だけ増減する。
- タプル長による制御：1タプル出力ポートに送る毎に、カウンタ値をタプル長分だけ増減する。

3　シミュレーションによる性能評価

3.1　シミュレーションモデル

シミュレーションでは、PMが8個ネットワークに結合され、タプル発生確率Pに基づいてタプルを発生させネットワークに入力する。それぞれのPMにある全タプル長を合計するとほぼ等しい値となっている。発生したタプルの属するバケットは128種類の中からランダムに決定される。シミュレーションではタプルがSUを通り抜ける時間をタプル長というが、このタプル長も発生時に決定される。タプル長の決定は、平均タプル長L_a、タプル長分布幅W_lとし、$[L_a-W_l, L_a+W_l]$の範囲からランダムに決定される。シミュレーションのモデルを図1に示す。シミュレーションでは、処理時間及びタプル数、タプル長合計についての平坦度を調べる。処理時間とはタプルの発生開始から全タプル転送終了までに要する時間のことである。平坦度は、平均標準偏差(MSD)という値を求めて調べる。D_{ij}をバケットiのPMjに格納されているサブバケットの大きさとすると、MSDは次のように求まる。

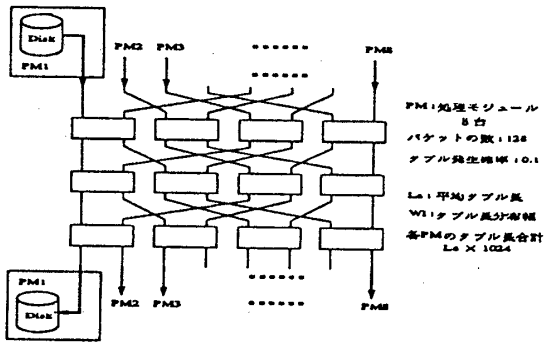

図1: シミュレーションモデル

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

ここで、$E\{\}$は期待値を表す。$MSD=0$が平坦分布となる。

3.2　シミュレーション結果

$(L_a, P)=(20, 0.05), (30, 0.1)$の場合について調べた。各PMに用意されているタプル長の合計は約$L_a \times 1024$とし、W_l及びしきい値Thrを変化させて調べた。

・タプル数による制御の場合：図2,3,4に結果を示す。横軸にはタプル長分布幅W_l、縦軸にはそれぞれ処理時間、タプル数のMSD、タプル長合計のMSDをとっている。

タプル長分布幅を大きくすると処理時間とタプル長合計のMSDも大きくなっているのが分かる。分布幅が大きいとネットワークに様々な長さのタプルが入力されるので、このことは予想できる。これに対し、タプル数のMSDは分布幅に関係なくほぼ一定の値となっている。このことから、数による制御では数の平坦化が分布幅に関係なく一定の効果を期待できるということが分かる。

・タプル長による制御の場合：図5,6,7に結果を示す。横軸にはタプル長分布幅W_l、縦軸にはそれぞれ処理時間、タプル数のMSD、タプル長合計のMSDをとっている。

この場合はタプル長分布幅を大きくすると処理時間とタプル数のMSDが大きくなっているのが分かる。これに対し、タプル長合計のMSDは分布幅に関係なくほぼ一定になっていることも分かる。タプル長による制御ではタプル長の合計を平坦化する際に一定の効果があるということが分かる。また、処理時間は数による制御と比べた場合に大きくなっていることも分かる。これは、カウンタの制御の仕方の違いから、タプルのSU通過率が低くなっていることを意味している。タプル数、タプル長合計のどちらについても平坦化後のMSDが小さくなっていることから、タプル長による制御も平坦化については有効であ

Algorithm of the Network with the Flat Bucket Distribution Mechanism for variable data length in the Super Database Computer (SDC)
Y.Aiba, S.Hirano, M.Kitsuregawa, M.Takagi
Institute of Industrial Science, University of Tokyo

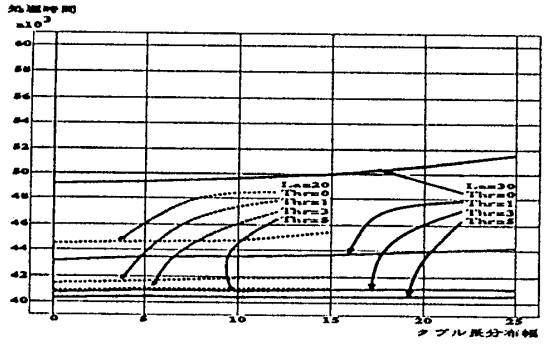

図 2: 処理時間とタブル長分布幅

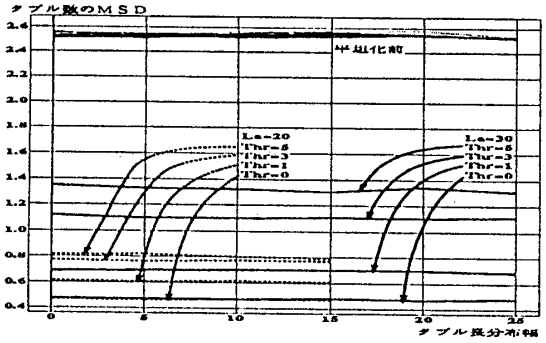

図 3: タブル数の MSD とタブル長分布幅

ることが分かった。

4 おわりに

本発表では、SDC の PM 間結合網に採用するバケット平坦化網の可変長データに対する制御方式とシミュレーションによる評価を報告した。シミュレーションによる性能評価の結果、本アルゴリズムが有効であることが分かった。

参考文献

[1] M. Kitsuregawa, S. Hirano, M. Harada, M. Nakamura, M. Takagi, "The Super Database Computer (SDC): 'Architecture, Algorithm and Preliminary Evaluation' ", *HICCS-25, 1992.*

[2] M. Kitsuregawa, Y. Ogawa, "Bucket Spreading Parallel Hash: A New, Robust, Parallel Hash Join Method for Data Skew in the Super Database Computer(SDC)", *the 16th Int. Conf. on VLDB, pp. 210-220, 1990.*

[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.

[4] 相場雄一、平野聡、喜連川優、高木幹雄 "スーパーデータベースコンピュータ用バケット平坦化オメガネットワークの非同期動作特性"、 第 42 回情報全国大会

[5] 相場雄一、平野聡、喜連川優、高木幹雄 "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オメガネットワークの動作特性"、 電子情報通信学会技術研究報告、 CPSY91-4-33

[6] 相場雄一、平野聡、喜連川優、高木幹雄 "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オメガネットワークの動作特性"、 第 43 回情報全国大会

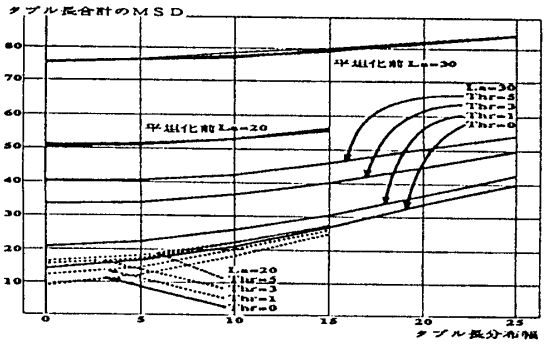

図 4: タブル長合計の MSD とタブル長分布幅

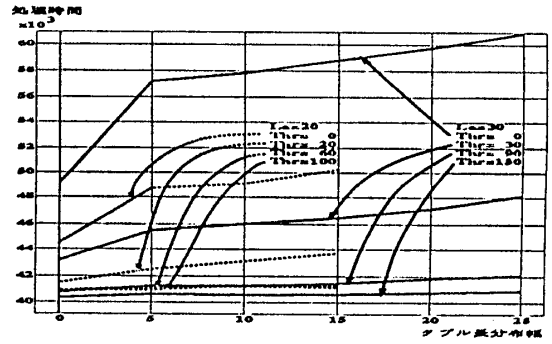

図 5: 処理時間とタブル長分布幅

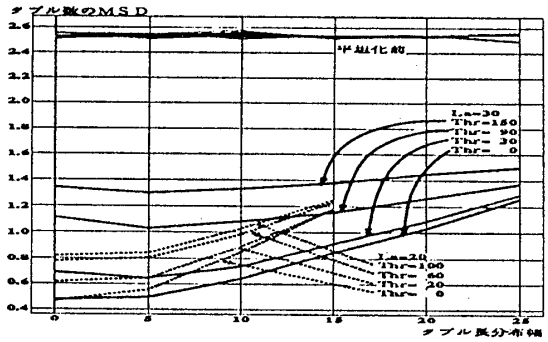

図 6: タブル数の平均標準偏差とタブル長分布幅

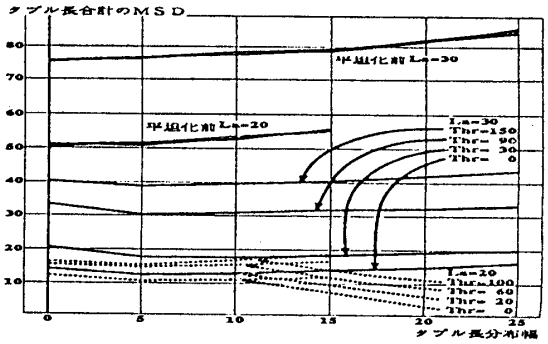

図 7: タブル長合計の平均標準偏差とタブル長分布幅