データベースマシン GRACE
=バケット送出機構=

4G-3

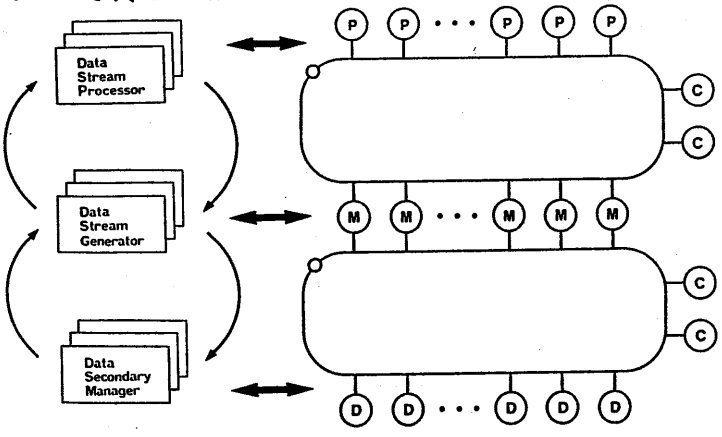
喜連川優　田中英彦　元岡達
東京大学　工学部

1. はじめに

Paula Hawthorn[1]によればデータベース処理は定型的操作を中心とする応用だけではなく、Joinが支配的な処理負荷の重い応用も存在し、これに適したデータベースマシンが望まれる。既に開発を行なって来たデータベースマシン[2][3]はHashとSortに基づく処理技法により、Join, Projection等処理負荷の重い関係代数演算をO(n)(n:メモリページサイズ)時間で高速処理する事ができ、Joinの多い応用に対しても高い性能を示すと考えられる。このマシンをGRACEと名付けたが、今回は、メモリモジュール(M.M)に於けるバケット送出機構を中心に述べる。

2. GRACE抽象アーキテクチャ

GRACEは図1に示される如く、3つの構成要素から成る。処理に際し、SDMからのデータ流は濾過されDSGにステージングされるが、この際、Join, Projectionアトリビュートに関してクラスタリングがなされる。その後DSGはバケットシリアルなデータ流を発生し、DSPはこの流れから適当なバケットを取り込みながら処理を行なうと共に結果タプルに対し、次演算に関するHashを施し、Hash Idを付加してDSGへ送出する。DSPの処理はHardware Sorterを用いる事により、O(n)時間で完了し、DSG群からの複数のデータ流にそってDSP群が処理を進めてゆく。DSGがバケット送出を開始し、結果が他のDSGに生成されるまでを1サイクルとし、これが1つの関係代数演算子(Join, Projection, Union)に対応する。このサイクルを繰返す事により問合せ木の処理が行なわれる。実際のマシンアーキテクチャでは図2に示される如く、各構成要素間は現在、リングバスによって結合することを考えている。以上の如く本マシンではデータ流発生器(DSG, MM)は単なるステージングバッファではなく主要な役割を果しており、次節でその構成について述べる。



P: Processing Module　C: Control Module
M: Memory Module　D: Disk Module

Fig.1 Abstract Architecture Of GRACE

Fig.2. Global Architecture Of Data Manipulation Subsystem In GRACE

3. メモリモジュールに於けるバケット送出機構

メモリモジュールの役割はプロセッシングモジュール群に対するデータ流の生成にある。M.Mはタプル入力時(ディスクからのステージング時及び、プロセッサ群による結果リレーションの生成時)には、その順序がランダムであったものを、出力時点でバケット順に変換する必要がある。

M.Mの記憶媒体としては半導体RAM、磁気バブル等が考えられる。前者ではディスクキャッシュや電子ディスクに見られる如く大容量化は着実に進歩しており、そのランダムアクセス性により前述の機能の実現は容易である。一方後者は前者に比べ容量の点で優れているが、アクセスタイム、転送レートはかなり劣る。しかし、今後DCデバイスによる速度の向上、CDデバイスによる容量の一層の増大が期待されている。RAMを用いてM.Mを構成する事も当然可能であるがここでは、磁気バブルの適用性について検討を行なった。

3.1 改良型M/mバブルチップ

● チップ構成……データベース処理では個々のレコードに対するアクセスタイムよりも寧ろ、レコード集合に対する実効データ

転送レコードが問題となる。M/mバブルでは不要レコードを1ビットタイムで読みとばす事が出来、転送レートの向上に有効であるが、容量の増大と共に、その効果が薄くなってしまう。そこでマイナループとメイジャライン間にバッファを設け、次に送出すべきレコードをバッファリングする事によって転送レートを大きく改善できると考えられる。このバッファはFIFOの場合、マイナループ長/マイナループ数×2段設ける事により、レコード間ギャップをなくせるが、実装上図3に示す如く、バッファループの付加が現実的となる。

● 制御……バブルにはその回転石磁界と同期したMark Bit RAMがブロック対応に付加されており、ここにはタプルが入力される時点で当該タプルのバケットIdが書き込まれる。タプル出力時にはループ間ゲートを制御し、当該バケットIdのタプル群をマイナループからバッファループへ順次移動する。

● 性能……図4にシミュレーションにより求めたチップの性能を示す。縦軸はギャップタイムのタプル転送時間に対する割合、横軸は送出対象レコードの全レコードに対する割合である。マイナループ長2048マイナループ数64とし、バッファループ長をパラメタとした。バッファループ長はマイナループ長に比べるとずっと小さい為ギャップタイムは大変少なくなる。バッファループは長いとバッファの効果は大きく、マイナループとの衝突は少ないが、逆に長すぎると再びバッファ内でのギャップが大きくなってしまう。図からわかる様に適当な長さのバッファを設ける事により、高い転送レートを確保出来る事がわかる。又バッファとマイナの対応は固定ではなく、少しずらせた方が良い特性の得られる事もわかる。

3.2 2バッファループチップ

図5はバッファループとメイジャラインの間にもう1つ小さなバッファを設け、更に性能の向上を図ったチップである。第0バッファはマイナループ長/マイナループ数×2より16とし、マイナループ一周分の余裕を持たせた。又この値は物理的大きさの制約にも関係している。第1バッファループはその大きさによる影響は比較的少なく、14、30でロードファクタ1/32以上で略1%以下の性能が得られた。今実効ブロック長を128ビットとすると、この値は1レコード当り1ビット程度のオーバヘッドとなり、略、限界特性と考えられる。これは回転をずらせた2つのバッファループを設ける事により、1つの場合に比べてマイナループとの衝突を大きく減少出来た事、及びループ0を短くする事でバッファ内ギャップを充分小さく出来た事によると考えられる。

3.3 バケット間連続処理

先の測定は最初のバケットに対する値であり、当該タプルはマイナループに存在すると仮定している。実際には次バケットの処理を重畳させ、当該バケットのデータと衝突しない限りバッファループに取り込む事が出来る。これにより、当該バケット出力終了時には次バケットのデータがバッファに充満している事になり、初期オーバヘッドは無視出来、一層性能を向上できる。

4. おわりに

バブルチップに僅かな改良を加える事により、必要なレコードだけを殆ど無駄なく連続して取り出せる事が示された。又制御は、マークビットを参照して衝突を判定しゲートの開閉を行なうだけであり、バブルの回転速度では容易に構成出来る。現在1バッファループ型のバブルチップによるMMの実装を検討中である。

参考文献 (1) P. Hawthorn "The Effect of Target Application on The Design of Data Base Machine" ACM SIGMOD 1981 (2) 喜連川他 信学技報 EC81-35 (3) 喜連川他 情報処理学23回全国大会 4F-5, 1981

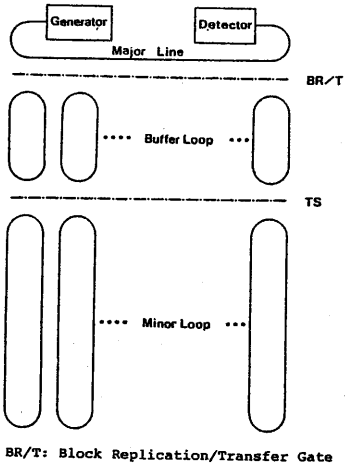


Fig. 3 Modified Major/minor Bubble Chip Organization

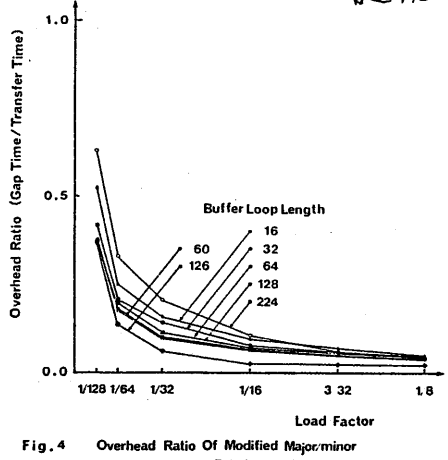


Fig. 4 Overhead Ratio Of Modified Major/minor Bubble Chip With One Buffer Loop

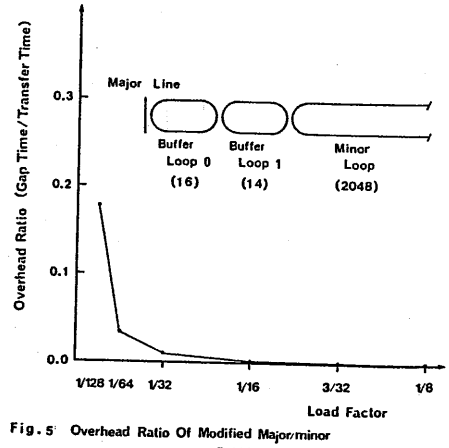


Fig. 5 Overhead Ratio Of Modified Major/minor Bubble Chip With Two Buffer Loops