2C-6　PIEにおけるモジュール間結合網の構成

坂井修一　丸山勉　田中英彦　元岡達
(東京大学　工学部)

1. はじめに

高並列推論エンジンPIEを開発中である。本マシンのような高並列の計算機システムでは通信系が処理の隘路になる可能性があり，転送性能の高い相互結合網の設計が必要となる。

本稿では，PIEの相互結合網とそれに関連する制御の方式に関して述べる。網設計の根拠として，本マシンの実行環境を想定して行ったシミュレーション結果(通信頻度，1回のデータ転送量など)を用いる。

2. PIEにおけるデータ転送と相互結合網

2.1 転送データの分類

PIEは，2レベルの階層構成をとる。レベル1(低レベル)システム内で転送されるデータは，大きくわけて5種類ある(表1。大きさ・転送頻度はシミュレーションの結果得られた値。T_uは目下80μs程度)。ゴールフレーム(GF)，コマンド(ノード情報)，構造データ，定義節，SMアドレスがこれである。レベル2の結合網では，GF，コマンドおよび定義節が転送される。

GF転送には高いスループットが，コマンドと構造データの転送には早いレスポンス(数μs)が要求される。定義節の転送頻度は他のデータのそれと比較して小さい。

2.2 相互結合網の構成と役割分担

PIEの相互結合網としては，上述の5種(厳密には7種)のデータそれぞれに独立の結合網を割り当てる構成が最も自然で，かつ高い転送性能を期待できる方法と考えられる。しかし，網のハードウェア量が大きく必要になる難点がある。そこで，大きさなどの点で共通点のあるGFと定義節は同一の網で，コマンド・構造データ・SMアドレスはそれぞれ独立な網で転送を行うことにする。

したがって，レベル1システムは4種類の網を持つことになる。分配網(DN)，コマンド網(CN)，追加読み出し網(LFN)，アドレス転送網(AN)がこれである(表2)。

同様にレベル2には，分配網(DN)とコマンド網(CN)がある。

表1. レベル1システム内の転送データ
Table 1. Transfer Data in Level 1 System.

データ		FROM	TO	大きさ	転送頻度*(/10)
GF		UP UP IM	MM IM MM	数100B	≦ 1
コマンド		AC AC AM	AC AM AC	10～20数B	2～3
構造データ	△G	UP	SM	10～20B	0.5 ～ 1
	LFコマンド	UP	SM	約10B	0.1 ～ 1.4
	LFデータ	SM	UP	10～50B	0.1 ～ 1.4
定義節		UP DSM	DSM DM	数100B 数100B	≪ 1 ≪ 1
SMアドレス		SM MM	UP SM	約4B 約4B	0.5 ～ 1

＊UP内での1GFの平均処理時間T_uを単位時間とした値

表2. PIEの相互結合網
Table 2. Interconnection Networks on PIE

結合網	転送データ	特　　徴
DN	GF 定義節	高スループット マルチキャスト
CN	コマンド	低遅延
LFN	△G LFコマンド LFデータ	はやいレスポンス (数μs以下)
AN	SMアドレス	△G生成に追随するスループット

3．各結合網の機能と実現方式

3.1 DN（レベル1）

DNでは，GFと定義節（ともに数100Bの大きさ）が転送される。このうちGFの転送においては，高スループット（1ポートあたり数MB/s）・均等な負荷分散・後のコマンド量を低く抑える，という3種の要求があり，定義節の転送においてはマルチキャスト機能が要求される場合がある。

ハードウェアのコストの点と要求される転送スループットの点から，レベル1のDNとして回線交換方式のオメガ網の適用を考えている。本オメガ網は，図1に示されるポート数4のスイッチング・ユニット（SU）を構成単位とする。SUは，網の閉塞を回避しつつ負荷の分散状況に適応した経路選択を行う機能，マルチキャスト機能などを持ち，ゲート数600余りで実現される。

3.2 CN・LFN・AN（レベル1）

CN，LFNはともに早いレスポンス（数μs）を要求される網であり，前者は最大15MB/s程度，後者は最大25MB/s程度のスループットが必要となる。両者とも集中管理型個別要求論理バスを適用する（バス幅はそれぞれ3B，5B程度）。

ANのトラヒックは約1.3MB/sである（転送遅延は問題にならない）。ANは，時分割多重チャネル方式のリングバスを用いて実現する（バス幅は1B程度）。

3.3 レベル2の結合網

レベル2のDN・CNに関しては，レベル1システム間でやりとりされるデータのトラヒックを見積り，その結果から実現方式の検討を行う予定である。DNとして3.1で述べた負荷分散適応型のオメガ網の適用，CNとして集中管理型個別要求論理バス（3．2）の適用が考えられる。

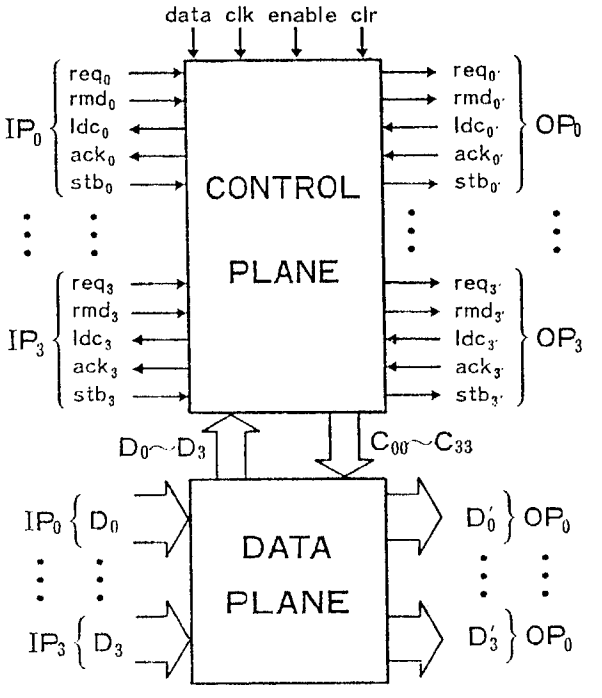

図1. 負荷分散適応型SU
Fig.1 SU Adaptive for Load Balancing.

4．考察

レベル1のDNとしてオメガ網を用いることを考えたが，これは1ルートの多段結合網であり，信頼性が低い欠点をもつ。SU内の論理回路の多重化，SUの付加による網の多ルート化などの対策を検討している。(1)

相互結合網を考慮に入れた負荷分散の評価（解析・シミュレーション）を行い，本稿で述べた方式の検証を行うこと，各網のインタフェース部の設計を行うことなどが今後の課題である。

5．おわりに

レベル1システムを中心に，PIEの相互結合網の構成に関して述べた。現在，網を組入れた階層的シミュレータにより評価を行っている。

文　献

(1)坂井，田中，元岡：“高並列推論エンジンPIEにおける相互結合網の構成”，信学技報，EC84-46.