2W−3

高速通信網 FEN (Fast Exclusive Network) による並列・分散処理環境

中條 拓伯†, 吉川 和宏†, 高橋 豊††, 前川 禎男†

† 神戸大学 工学部 情報知能工学科　†† 近畿大学 理工学部 経営工学科

1．はじめに

現在，単一プロセッサの処理能力の限界から，並列処理に期待が寄せられ，種々のシステムが商品化され，稼働している．しかしながら，並列処理アルゴリズムや並列分散OSなど，ソフトウェアに関する研究が追随せず，並列処理システムの真価を充分に発揮するには至っていない．その理由として，並列計算機自身が一般には普及しておらず，多数のエンドユーザが並列処理プログラミングを行なえる環境にはなく，教育現場において並列処理アルゴリズムに関する充分な教育が行なえないという点が指摘されている．

一方，ワークステーション(WS)の低価格化とネットワーク環境の充実により，LANにより接続されたWS群を利用した分散処理に関する研究が盛んに行なわれている．しかしながら，標準化されたイーサネットは，ファイルなどの比較的大きいサイズのデータに対しては有効であるが，並列処理における同期処理などに使用される細かなデータ転送に対しては速度的に不十分である．

そこで，我々は研究室内などの限られた範囲に設置されたWS群を高速の専用シリアル通信網により接続し，身近なWSを利用したポータブルな並列処理環境の構築を目的としたシステムの開発を進めている．専用のシリアル通信網をFEN (Fast Exclusive Network) と呼ぶ．

ここでは，まずFENにより接続されたシステムの全体構成について述べる．次に，FENのハードウェア／ソフトウェア・インタフェースについて説明する．そして，現在の設計段階における，WSノード間の通信性能予測を示し，最後に現状と今後について述べる．

2．システム構成

図1にシステムの全体構成を示す．FENハードウェア・インタフェース(FEN-HI)は，ポータビリティを確保するためにWSの汎用入出力インタフェース(SCSIなど)に接続される．インタフェースは複数のシリアルポートを持ち，隣接間WSをメッシュやハイパーキューブといったトポロジーで相互接続する．シリアル通信には，パラレルーシリアル変換によるレイテンシなど速度的な問題はある．しかし，実装上の利点から，多数のポートを設置して，高次元のネットワークを構成することにより，転送時におけるノードのホップ数を抑えることができる．そして，中継するノードにおけるバッファをある程度確保し，スイッチングには，転送処理をパイプライン化するVirtual Cut-through方式[1]を採用することにより，ノードにおける中継処理のオーバヘッドを抑える．さらに，ルーティングなどの転送処理をハードウェア化することによって，ネットワークにおけるデータ転送のバンド幅の拡大をはかる．

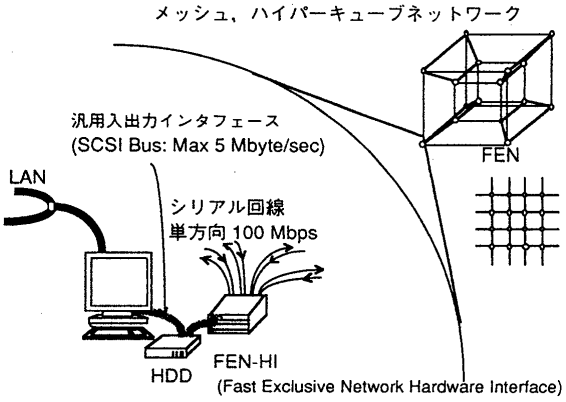

図1 システムの全体構成

FENソフトウェア・インタフェース(FEN-SI)はユーザに，分散環境における様々な透過性を与える．各ノードではネットワークに接続されたWSの台数を意識することなく処理できる．また，ネットワークのトポロジーはユーザには見えず，仮想的に共有メモリを持たせることも可能である．

3．FENインタフェース

3．1　ハードウェア・インタフェース

現在，設計を進めているFENハードウェア・インタフェースの構成を図2に示す．FEN-HIとWSとの間の入出力インタフェースにはSCSIを採用した．SCSIバスは現在多くのWSに採用されており，対象機種のデバイスドライバを用意することにより各機種に対応可能でき，異機種間のネットワークも構築できる．

FEN-HIは，(1) SCSIコントローラとバッファメモリブロック，(2) CPUブロック，そして(3) 通信ブロックの3ブロックに大別される．ブロック(1)に使用するSCSIコントローラはSCSI 2規格対応であり，最大5 Mbyte/secの転送速度を持つ．デバイスドライバから指令を受けたコントローラはバッファメモリを介して，WSとデータの授受を行なう．また，バッファメモリは仮想共有メモリを構築する際には，共有空間へアクセスするときのキャッシュとして働く．

CPUブロックのCPUは，おもにメッセージパケットのルーティング処理や，SCSIコントローラのコマンド処理などを行なう．また，仮想共有メモリを実現する際には，データの無矛盾性を保証するためのコンシステンシ・コントロールも行なう．以上の処理は高速性を要求されるが，種々のルーティングアルゴリズムやコンシステンシ・プロトコルを開発するためにはプ

Parallel and Distributed Processing Environment through the Fast Exclusive Network (FEN)

Hironori NAKAJO†, Kazuhiro YOSHIKAWA†, Yutaka TAKAHASHI†† and Sadao MAEKAWA†　e-mail: nakajo@seg.kobe-u.ac.jp

† *Faculty of Engineering, Kobe University*　†† *The Faculty of Science and Technology, Kinki University*

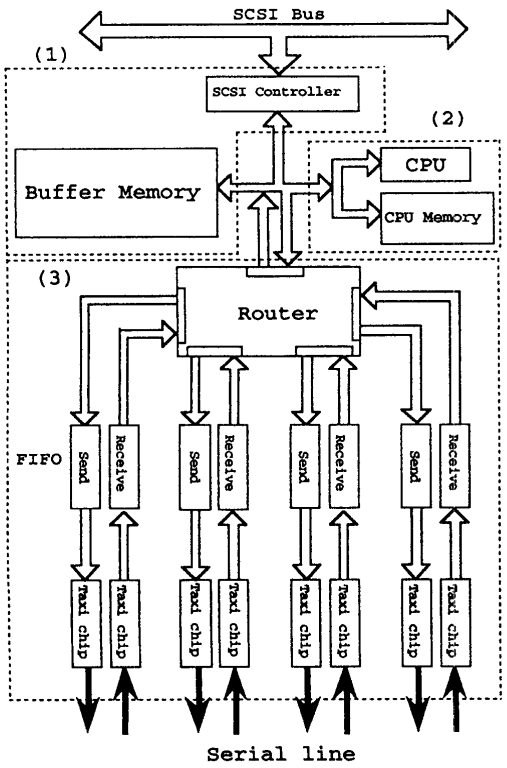

図2 FEN ハードウェア・インタフェース

ログラマブルである必要がある．しかし，汎用のCISCチップでは充分な処理速度は得られず，逆にRISCチップを用いた場合は実装上において困難である．以上の点を考慮して，1命令を1クロックで実行でき，高級言語(C言語)利用できるモトローラ社製のDSP MC56001をコントロールCPUとして採用した．

通信ブロックには4つの送受信シリアルポートを持つ．隣接ノード間におけるポート間の通信にはAMD社製のTaxi chipを用いる．Taxi chipは送信用(トランスミッタ)と受信用(レシーバ)があり，その間を高速に(Max 100 Mbps)シリアル通信することができる．一般的にシリアル通信では転送先のバッファ(FIFOなど)の状態を把握するために複雑なプロトコルが要求され，通信効率が抑制される．しかし，FEN-HIではTaxi chip間においてデータ転送の可否を示すメッセージをほぼ瞬時に交換し，安全かつ効率のよいデータ転送を可能にしている．また，ノード上を通過するデータを効率的に中継するためにルータを装備し，FIFOに送り込まれたパケットが自ノード宛でなければ，ルータによりそのまま隣接ノードに転送される．転送のスイッチングにはVirtual Cut-through方式を採用する．したがって，数パケットを格納できるFIFOを装備することにより，転送経路上のトラフィックが増えた場合はStore-and-Forward方式のようなスイッチング方式となり，転送経路上のパケットが他のパケット転送の妨げとなることを防止する．

3．2　ソフトウェア・インタフェース

FEN-HIはWSからはSCSIに接続されたインテリジェントデバイスとして認識される．FEN-HIにアクセスするためには，専用のデバイスドライバを用意し，このデバイスドライバを介することによって，ユーザに対して統一的なインタフェースを提供する．

表1　FEN-HIの各部仕様

WS - FEN-HI間の転送速度	5Mbyte/s
FEN-HI内におけるパケットヘッダの解析およびルーティング処理時間	20〜50 μsec (平均35 μsec)
Taxi chip間の転送速度	100Mbps

FENソフトウェア・インタフェース(FEN-SI)では，プログラミング環境において2種類のインタフェースをユーザに提供する．1つはノード間で高速にデータ通信を行なうメッセージパッシング・インタフェースである．この場合，ユーザは転送先のノードIDを含んだパケットを陽に作成し，FEN-SIに転送する．ここで，CPUブロックのDSPは，パケットのルーティングのみを行ない，ノード間の高速転送を可能とする．

さらに，FEN-SIは仮想共有・インタフェースを提供する．この場合，ユーザから見たFEN-SIは，他のノードと仮想的に共有している空間に見える．ユーザは共有空間に対するリード／ライトによって通信を行ない，密結合並列計算機における並列プログラミング・パラダイムが提供される．共有空間はページ単位に分割され，DSPによりオーナシップに基づいたコンシステンシ・プロトコルにより管理される．この場合，プロトコルに用いられるメッセージパケットはDSPが作成する．

4．FEN性能予測

現在の設計に基づいたシステムの仕様を表1に示す．この仕様に基づき，1 Kbyteのページデータの転送速度について性能予測を行なう．WSからFEN-HI内のバッファメモリへの転送時間は200 μsec，次に，インタフェース上のDSPにおけるパケットヘッダ解析処理とルーティング処理時間を20 μsec ～50 μsecと見積る[2]．隣接Taxi chip間では80 μsecの転送時間を要する．DSPおけるヘッダ解析処理と隣接Taxi chip間データ転送処理はVirtual Cut-through方式により，オーバラップして行なわれ，ヘッダ部分の転送時間はヘッダ解析処理に比べて微小であり，無視できる．以上の転送処理時間Tを，通信するノード間距離をN (N >= 2)として定式化すると

$$T = 200 + 35 + 80 + 35 \times (N - 1) + 200 = 480 + 35 \times N \ (\mu\text{sec})$$

として求められる．FENのネットワークトポロジーを8 x 8のトーラスネットワークと仮定した場合，最も長いノード間距離は8ノードとなり，約760 μsecで転送できる．

5．現状と今後について

現在，FEN-HIの詳細設計を終え，通信ブロックについては論理シミュレータにより動作確認を行なった．今後はFEN-HIを完成させた後，UNIXからアクセスするためのデバイスドライバを作成する．FEN-SIについては，メッセージパッシング・インタフェースにおける通信プロトコル，および仮想共有・インタフェースにおけるコンシステンシ・プロトコルの詳細について現在検討中である．

参考文献

[1] Kermani, P. and Kleinrok, L. : Virtual Cut-Through: A New Computer Communication Switching Technique, Computer Networks, vol.3, pp.267-286 (1979)
[1] 中條 他：リング結合型並列計算機 KORP における分散共有メモリシステム，プロトタイプの性能評価，並列処理シンポジウム JSPP'92, pp.203-210 (1992)
[2] 鈴木洋：広帯域ISDNを用いた高速コンピュータネットワーク技術動向，コンピュートロール特集No.38／コンピュータネットワーク，pp.9-15 (1992)