Manhattan Street Network における代理送受信を用いたルーティング手法

江草俊文*・小畑正貴**

*岡山理科大学工学研究科博士課程システム科学専攻

**岡山理科大学工学部情報工学科

(1998年10月5日　受理)

1. はじめに

ローカルエリア・メトロポリタンエリアのパケット通信を行なうために設計されたManhattan Street Network (MSN) [1]は，単方向の通信メディアをトーラス状に接続し，通信可能な方向を互い違いにすることで単方向の欠点をカバーしたネットワークである。

単方向化することで，双方向のトーラスネットワークよりも少ないリソースで構成できるなどの利点がある。また，より単純なハードウェアを用いることでルーターの動作周波数をあげることが容易になり，その結果として複雑なハードウェアを持ったネットワークよりも良い性能が得られる場合もある[2]。

しかし，従来のMSNのルーティング法[3]では

- デッドロックへの対応がなされていない
- メッセージのFIFO性が保証されない
- ルーティングアルゴリズムが複雑である
- 効率的なブロードキャストが出来ない

などの多くの問題点があり，そのまま並列計算機のネットワークとして用いることは難しい。

そこで我々は，MSNを並列計算機のネットワークへ応用するために代理送受信ノードの概念を導入し，

- デッドロックフリーの保証
- メッセージのFIFO性の保証
- 中断ノードでの容易なルーティング
- 効率的なブロードキャストが可能

などの特徴を持ったManhattan Street Network with Proxy Send/Receive (MSN/P)を提案する。

2章では，我々の提案するMSN/Pの概要，ルーティングアルゴリズム，およびデッドロック回避法について詳述し，3章ではソフトウェアシミュレーションによる性能評価について述べる。4章では，まとめとMSN/Pの3次元への拡張，単方向ネットワークのメッ

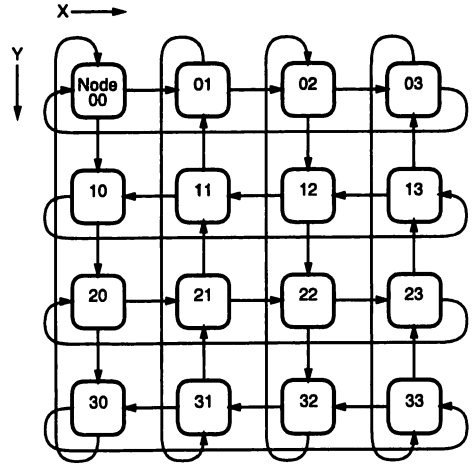

図1 MSN with Proxd/Receive のトポロジ

シュへの応用について簡単に紹介する。

2. MSN with Proxy Send/Receive

2.1 MSN with Proxy Send/Receive の概要

並列計算機のネットワークでは，デッドロック対策が必須である。また，メッセージのFIFO 性を保証することで，通信プロトコルの単純化が可能となる。また，受信メッセージの再構築が不要となる。

そこで我々は，MSN のネットワークトポロジ(図1)をそのまま用い，従来のルーティングに化理送受信の概念を採り入れることでデットロックフリー，メッセージの FIFO 性の保証，効率的なブロードキャスト，容易な中断ノードでのルーティング制御が実現できる Manhattan Street Network with Proxy Send/Receive (MSN/P) を提案する。

2.2 MSN/P のルーティング

MSN/P でのルーティングは，Xリンクを優先して使い，その後Yリンクを用いて受信ノードへと転送する次元順ルーティングを行なう。ただし，MSN/P では通信可能なリンクの向きが強く制限されているので，単純にXリンクを用いると受信ノードから遠ざかってしまう場合がある。そのため，代理送信ノード・代理受信ノードという概念を導入することで，出来るだけ最短に近い経路を選択できるようにする。

(1) 必要があれば，代理送信ノードへ送信する。

(2) Xリンクを用いて，受信ノードに近付く。

(3) Yリンクを用いて，受信ノードに近付く。

(4) 必要があれば，代理受信ノードから受信ノードへ送信する。

代理送信・代理受信ノードを用いることで，その間のルーティングは，単純なXリンク優

先になり，かつ，大抵の場合は最短経路が選択される。

MSN/P では，まず，送信開始時にルーティング情報の生成を行なう。ルーティング情報は，代理送信・受信のフラグおよびX・Yの相対アドレスで構成される。代理送信・受信のフラグは，真なら代理送信・受信が必要であることを示す。相対アドレスは，送信ノードと受信ノードの絶対アドレスの差の絶対値である。しかし，送信・受信に代理を用いる場合は，代理ノードでそれぞそ送信・受信ノードと読み替えて相対アドレスを求める。

次に実際のルーティングであるが，送信ノードでは代理送信が必要ならばYリンクを用いて代理送信を行ない，不要ならば中継ノードと同様の処理を行なう。中継ノードでは，Xリンクを優先して送信し，その後Yリンクを使って送信をする。その際，X・Yの相対アドレスをデクリメントする。相対アドレスが（0，0）になれば，受信ノードに達したとして判断されるが，代理受信フラグが真の場合は，Xリンクを用いて1回だけ送信をする。

本手法と従来の MSN のルーティングアルゴリズムの違いは，以下である。

- デッドロックフリーが保証される
- メッセージの FIFO 性が保証される
- 効率良くブロードキャスト出来る
- 送信ノードでのルーティング情報生成は MSN のルーティングアルゴリズムと同程度に複雑だが，中継ノードでの判定は，MSN/P が非常に単純でハードウェアでの実装も容易

図2に代理送受信が必要となる場合の通信パターンの例を示す。MSN/P では，送信ノードの通信可能なリンクの向きによって，4通りのマップが出来る。これは，黒く塗りつぶされたノードを送信ノードとし，グレーで塗りつぶされたノードは代理受信が必要となるノード，代理受信の補正後に矩形で囲まれた領域内にあるノードは，代理送信が必要となるノードである。また，送信ノードは代理送信が必要なノードに含まれる。これは，ブロードキャスト時にメッセージのFIFO性を保証するために必要である。図3は実際のルーティングの例である。ノードSから，ノードD0，D1，D2，D3に送信した場合，それぞれ，

D0 代理送信ノードも代理受信ノードも不要な場合。Xリンク優先でルーティングされる。

D1 代理受信ノード（D1'）が必要な場合，SからD1'は通常のルーティング，D1'からD1は代理チャネルを用いてルーティングを行なう。

D2 代理送信ノード（S'）が必要な場合，SからS'は代理チャネルを用いてルーティングを行ない，S'からD2への通常のルーティングを行なう。

D3 代理送信ノード（S'），代理受信ノード（D3'）が必要な場合。

2.3 デッドロック回避

デッドロックの回避については，ネットワークの物理的な循環構造を仮想チャネルを用いて論理的に断ち切る方法を採用する[4]。この方法は，仮想チャネルの構成に必要になる

バッファの量を最小に押えることができる。

具体的には，次に示す優先順位と条件でチャネル変便を行なう[5),6)]（図3）。

(1) 代理受信ノードから受信ノードへの送信は，代理チャネル（X）

(2) ラップアラウンドを通過する場合にはチャネル1

(3) 方向転換をする場合は，チャネル0

(4) 代理送信ノードへの送信は，代理チャネル（Y）

したがって，Yリンク，Xリンクに対して，それぞれ3チャネルの仮想チャネルを設けることで論理的な循環構造がなくなるので，デットロックしない。図4はMSN/Pのチャネル依存グラフである。この図において，YPは代理チャネル（Y），X0/1はそれぞれXチャネル0/1，Y0/1はそれぞれYチャネル0/1，XPは代理チャネル（X）を示す。

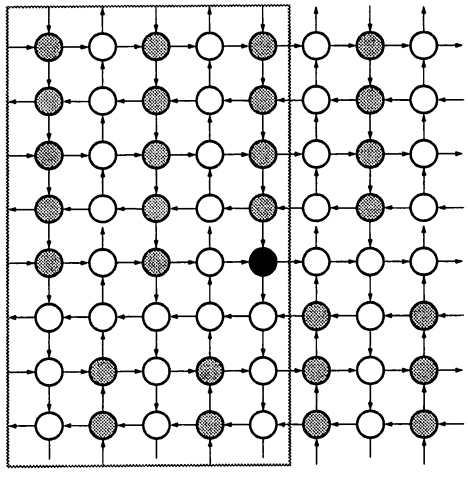

図2 代理送受信フラグのマップ

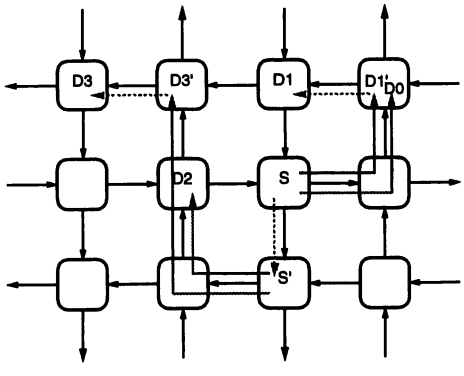

図3 ルーティングの例

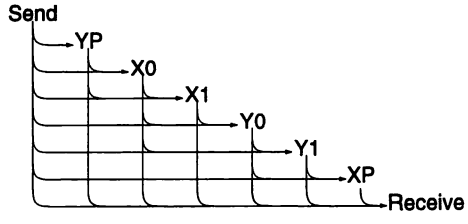

図4 MSN/Pのチャネル依存グラフ

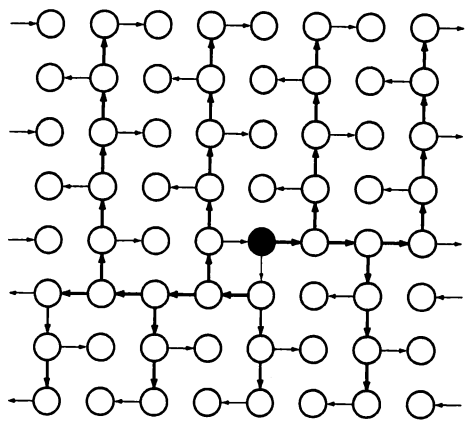

図5　ブロードキャストの一例

2．4　ブロードキャスト

MSN/P ではブロードキャストセルに対して，自プロセッサ，Xリンク，Yリンクの3方向に対して通常のルーティングと同じ規則でセルを送信すれば良い。また，$n \times m$ の MSN/P では，送信元を（0，0）と見なし相対アドレスが，$(-n/2, -m/2)$ — $(n/2-1, m/2-1)$の矩形からはみ出さないようにブロードキャストの終了条件を設定すれば良い。図5は，黒く塗りつぶされたノードを送信ノードとした場合に，ブロードキャストで選択される経路が矢印で示されている。細い矢印は代理送受信での通信を示している。図5からも明らかなように $(-n/2, -m/2)$ — $(n/2-1, m/2-1)$ の矩形からはみ出すような通信は，すべて代理受信ノードからの送信となるため，相対アドレスは矩形から出ることはない。

また，Point-To-Point の通信の場合と同じ経路を通るので，

- デッドロックフリー
- メッセージの FIFO 性

が常に保証されていることは明らかである。

3．ソフトウェアシュミレーションによる評価

3．1　シミュレーションの条件

単方向トーラスネットワークと双方向トーラスネットワークの通信性能を比較するために，通信シミュレーターを作成し実験を行なった。

シミュレーションの対象は，

Bi-dir 双方向2次元トーラスネットワーク。e-cube ルーティングを用いる。半二重で通信を行なう。

MSN 文献[3]で紹介されている最短経路を選択する Rule 1 を用い，セル衝突時は迂回経

路を選択する動的ルーティングを行なう。このため他の方法と異なり，メッセージのFIFO 性はない。

MSN/P ルーティルグや，デッドロック回避については，先に述べた方法を用いる。

Simple 通信可能な方向がすべて同じ単方向トーラスネットワーク。次元順ルーティングを用いる。

シミュレーションの条件として

(1) 各ノードは，異なる方向の送受信を同時に行なえる。

(2) 1セルの通信に必要な時間を1単位時間とする。

(3) 簡単のため Store & Forward を用いる。

(4) ノード構成は，図6に示したものを用いた。

MSN X，Y共に深さ4の FIFO

MSN/P X，Yの仮想チャネル0を深さ2の FIFO

Simple X，Yのすべての仮想チャネルを深さ2の FIFO

(5) 通信の形態は，

Random 不規則で局所性を持たない通信

Nearest Neighbor 隣接ノードの中から，ランダムに選択される1つに対して行なわれる通信

Reduce すべてのセルが，（0，0）に集中する通信

Hot Spot 発生するセルのうち50%が（2，j）を受信ノードとする通信の4種類について。

(6) 通信時間は，時間0でセルが発生し，それが受信ノードに到達するまでの時間とし，途中でセルを生成しない。

(7) スループットは，ネットワーク上に常に同数のセルが存在するようにし，対象とするすべてのネットワークが定常状態になった時間600から，時間1000の400単位時間に到達した平均のセル教とした。

ノード構成を図6の様に設定した理由は，ハードウエアコストをできるだけ近付けるためで，すべてのネットワークに関してノードあたりのバッファ数が9になるように，FIFOを追加したためである。また，FIFO を追加する場所については，FIFO がない場合のネットワークをシミュレーションし，通信路の衝突頻度が高かった場所に設定した。シミュレーションは，8×8ノードの場合について行ない，その結果を図7から図14に示す。通信時間は，横軸に時間0で発生させるセル数，縦軸に通信に要した時間をとった。またスループットは，横軸にネットワーク上に存在するセル数，縦軸にスループットをとった。

3.2 シミュレーションの結果

全てのシミュレーションの結果は，Simple が悪く，Bi-Dir がよい。MSN，MSN/P は Simple と Bi-Dir の中間の結果が得られた。

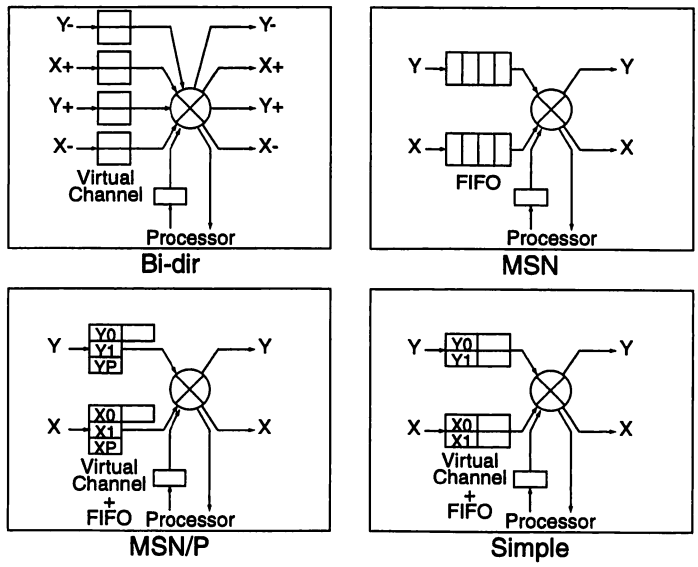

図6　ノード構成

"Random" の平均通信時間（図7） MSN が図1のように，通信可能な方向を互い違いに配置したために，Bi-dir と比較して Simple 程ネットワークの直径が大きくなっていないことを示している。また，MSN では動的ルーティングを行なっているが，動的ルーティングを行なわない MSN/P と大差のない結果が得られた。また，MSN 以外はセル密度があがるにつれて通信時間も伸びているが，MSN はほぼ一定である。

"Random" のスループット（図8） Bi-dir が一番良いが，MSN，MSN/P ともに Bi-dir に対して2割程度の悪化に留まる。MSN，MSN/P はほぼ同等のスループットが得られる。

"Nearest Neighbor" の平均通信時間（図9） Bi-dir では平均通信時間が若干増加傾向なのに対して，MSN，MSN/P ではセル密度に関わらずほぼ一定になっている。これは，Bi-dir では，隣接ノードから同時に送信されたセルとの衝突により通信待ちがおこるが，MSN，MSN/P ではおこらないためである。

"Nearest Neighbor" のスループット（図10） Simple では，Bi-dir の3割程度まで性能が落ち込むが，MSN，MSN/P では，Bi-dir の6割程度の性能を保っている。

"Reduce" の平均通信時間（図11） 通信バッファープロセッサ間も1単位時間で1セルしか受けとらないため，全てのネットワークでほぼ同じ結果が得られた。

"Reduce"のスループット（図12） 通信バッファープロセッサ間も1単位時間で1セルしか受けとらないため，全のネットワークでほぼ同じ結果が得られた。

"Hot Spot" の平均通信時間（図13） セル密度が高くなると，すべてのネットワークについて性能が悪くなっていくが，MSN は比較的悪化の度合が少ない。MSN/P は Bi-dir

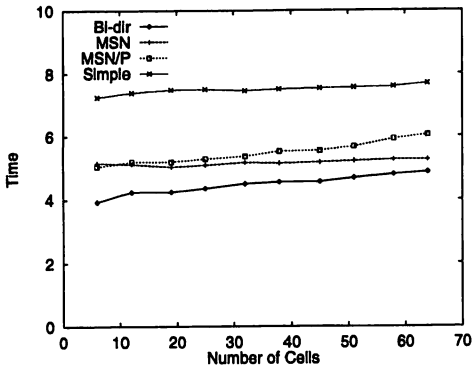

図7 “Random” の平均通信時間

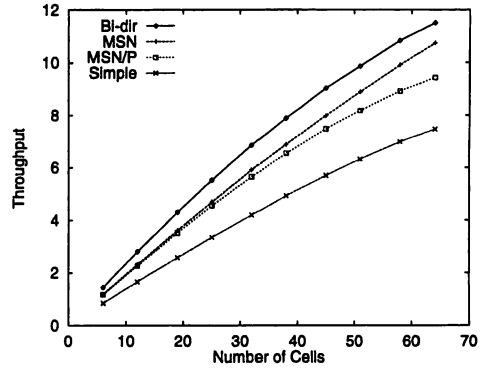

図8 “Random” のスループット

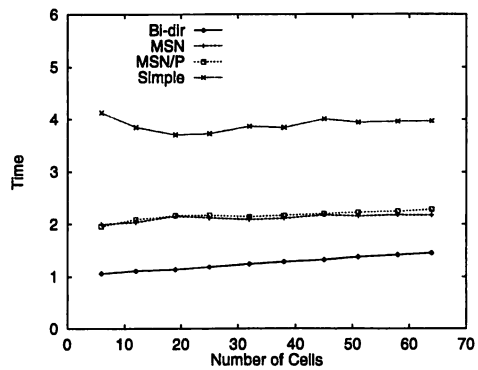

図9 “Nearest Neighbor” の平均通信時間

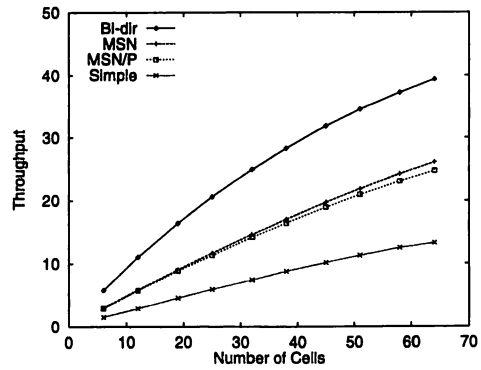

図10 “Nearest Neighbor” のスループット

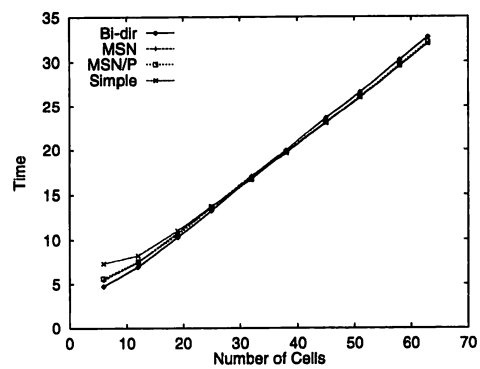

図11 “Reduce” の平均通信時間

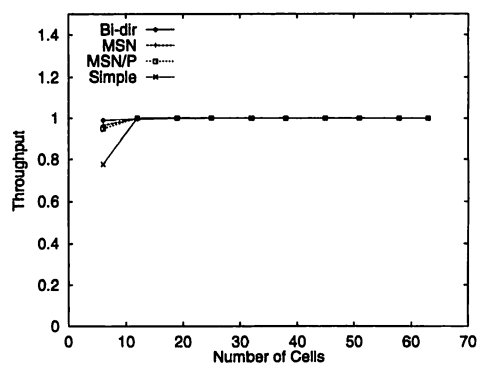

図12 “Reduce” のスループット

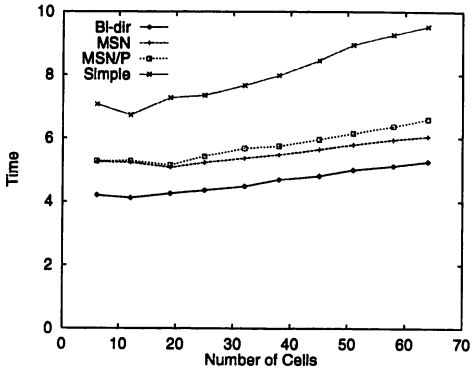

図13　“Hot Spot” の平均通信時間

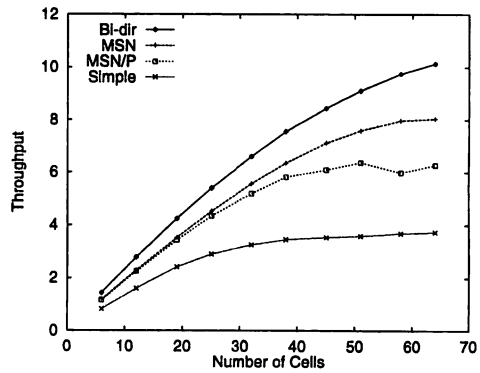

図14　“Hot Spot” のスループット

とほぼ並行して悪化していく。

“Hot Spot” のスループット（図14） MSN，MSN/P ともにセル密度が上がると性能低下が目立ち，Bi-Dir と比較して，それぞれ 8 割，6 割程度まで落ちている。Bi-Dir より経路選択の制限が強いため，セルの衝突が頻繁に起こっているためである。また，MSN よりもさらに経路選択の制限が強い MSN/P ではいっそう目立っている。

ここでは結果を示していないが，32×32ノードでのシミュレーションも行なった。その結果から，ノード数が多くなった場合，平均通信時間，スループットともに Simple は極端に性能が悪化するのに対して，MSN，MSN/P はより Bi-Dir に近い性能を示すようになる。

また，MSN は “Hot-Spot” のシミュレーションではデッドロックを起こし正しい計測が出来なくなってしまった。Bi-dir，MSN/P，Simple ではデッドロックはもちろん起きない。

4．ま と め

MSN/P の概要，ルーティングアルゴリズム，デットロック回避法について述べた。MSN/P は，MSN のトポロジを用い，代理送信・受信ノードの概念を導入した新たなルーティングアルゴリズムを用いることで MSN の欠点を解消し，

- デッドロックフリーの保証
- メッセージのFIFO 性の保証
- ブロードキャストへの対応
- 中継ノードでのルーティングアルゴリズムの簡素化

という特徴を持たせることが出来る。

また，ソフトウエアシミュレーションによる通信性能の比較によって，MSN/P はほぼ MSN と同様の性能であることを示し，特にノード数が多い場合のランダム通信において

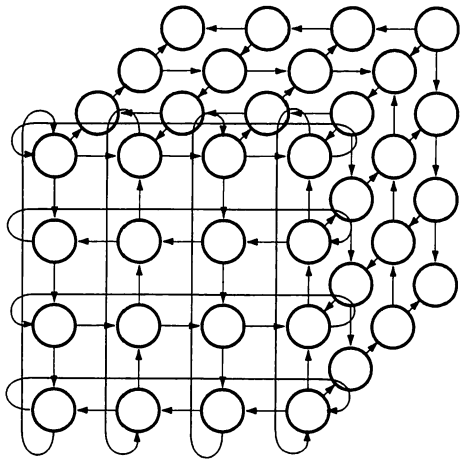

図15 MSN/P-3D

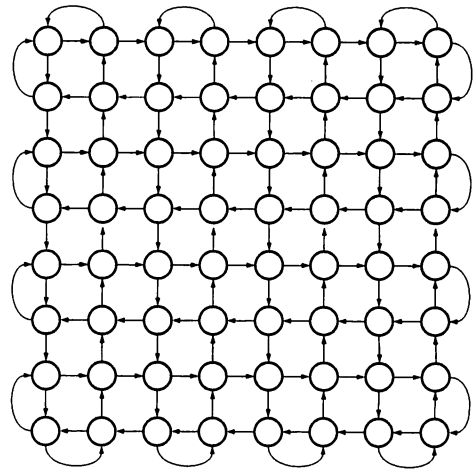

図16 MSN/P-Mesh

は，双方向トーラスとほぼ同様の性能であることを示した。

MSN/P で用いた代理送受信を用いることで，MSN を3次元に拡張することが出来る（図15）。このネットワークでは，3入力3出力のリンクが必要になり，X/Y/Zのリンクに対してそれぞれ3チャネルの仮想チャネルを持たせ，各リンクに対して代理ノードを設定することで，デットロックフリーなど MSN/P の特徴を保つことが出来る。

また，代理受信ノードを2つにすることで，単方向メッシュ（図16）のルーティングも行なえるのうになる。このネットワークでは，代理送受信が必要かどうかの判定も複雑になるが，すべて配線が平面上で行なえるという利点がある。One-Chip Multi Processor のように3次元的な配線を行ないにくいものや，計算機実習室などで WS・PC クラスタを構成する場合のように計算機が空間的に広く配置されラップアラウンドの配線が困難な場合などへの応用が考えられる。

参考文献

1）N. F. Maxemchuk：The Manhattan street network, *Proc. GLOBECOM* '**85**, pp. 255—261 (1986).

2）林 匡哉，堀田真貴，大津金光，吉永 努，馬場敬信：HDL 設計に基づく並列計算機ルータの評価，情報処理学会研究報告98-ARC-130, Vol. **98**, No. **70**, pp. 79—84 (1998).

3）N. F. Maxemchuk：Routing in the Manhattan Street Network, *IEEE Transactions on Communications*, Vol. COM-**35**, No. **5**, pp. 503—512 (1987).

4）天野英晴：並列コンピュータ，昭晃堂 (1996).

5）江草俊文，小畑正貴，単方向2次元トーラスネットワークの構成と，シミュレーションによる評価，情報処理学会研究報告96-ARC-120, Vol. **96**, No. **106**, pp. 13—18 (1996).

6）可児純一，江草俊文，小畑正貴，PC クラスタのための単方向2次元トーラス網用ルータ，電気・情報関連学会中国支部第48回連合大会講演論文集，p. 449 (1997).

Manhattan Street Network with Proxy Send/Receive

Toshifumi EGUSA* andMasaki KOHATA**
*The Graduate School of Engineering,
Okayama University of Science
**Dept. of Information and Computer Engineering,
Faculty of Engineering,
Okayama University of Science
(Received October 5, 1998)

The Manhattan Street. Network (MSN) is a network technology designed for packet communications in a local or metropolitan area. In this paper, we propose a new routing algorithm, MSN/P, based on "Proxy Send/Receive" model for the MSN. The new routing algorithm is simpler than the original one, and has the following features, that is, deadlock free, FIFO order message transfer, broadcast and so on.

We evaluate the MSN/P by comparison with the MSN and uni- or bi-directional torus in network performance. As the results, it is shown that the performance of MSN/P is comparable to MSN and the MSN/P offers the various advantages of inter-connection network of parallel computer.