計算機アーキテクチャ 124－7
（1997. 6. 27）

ウェーハスタック構造型・超並列コンピュータネットワーク:TESH

堀口 進[†]　　大木 孝之[††]　　V. Jain[†††]

[†]北陸先端科学技術大学院大学 情報科学研究科
[††]現、日本ビクター株式会社
[†††] University of South Florida

近年のVLSI技術の進歩によって、超並列計算機システムを1枚のウェーハ上に実装するWafer Scale Integration(WSI)が実現可能になってきた。しかし、1枚のウェーハ上に実装できるプロセッシングエレメント(PE)数は限られているため、より大規模な超並列計算機システムを実現する新しいWSIの手法として、ウェーハを立体的に積み重ねたウェーハスタック構造が提案されている。ウェーハスタック構造を超並列システムをウェーハスタック構造を用いて実現する場合、ウェーハ間結線に必要な幅が数百μに達するため、ウェーハ間結線を減少させる必要がある。しかし結線数を少なくすると、直径やプロセッサ間平均距離が増加し、著しい性能低下が起こる。また冗長リンクが少なくなりフォールトトレランス性能が悪くなる。従って、ウェーハ間の結線を少なく保ち、ネットワーク全体の特性を上げる必要がある。本研究ではウェーハスタック構造を用いて超並列システムを実現する3次元階層型ネットワークTESHについて、ウェーハ間結線数、直径、レイアウト面積を用いて詳しく検討する。

キーワード：超並列計算機システム、WSI、ウェーハスタック構造、ウェーハ間結線、相互結合網

Interconnection Network TESH for Massively Parallel Computers on Stacked Wafers

Susumu Horiguchi[†]　　Takayuki Ooki[††]　　V. Jain[†††]

[†] Graduate School of Information Science ,JAIST
[††] Japan Victor Corp.
[†††] University of South Florida, U.S.A.

A massively parallel computer, 3D-computer was implemented in wafer scale integration(WSI). However the number of processing elements(PEs) implemented on a wafer is limited. The three dimensional(3D) stacked implementation has been proposed as a new technology for massively parallel computers. Since the area of vertical links between wafers amount to hundreds μm^2, it requires far fewer number of vertical links than almost all known massively parallel computer networks. A restriction of vertical links increases the diameter of networks and makes network feature worse. Thus, a massively parallel interconnection requires reducing the number of vertical links for 3D stacked implementation by keeping good network feature.

This paper discusses network feature such as diameter, the maximum number of vertical links and chip area of conventional networks TESH(Tri connected mESHes) in 3D stacked implementation.

keywords: Massively parallel computer, Wafer scale integration(WSI), 3D stacked implementation, Interconnection networks

1 はじめに

超並列コンピュータの相互結合網をウェーハスタック構造上 [1][2] にマッピングする場合、ネットワークを複数のウェーハに分割配置する必要がある。このため、ウェーハスタック構造ではウェーハ間結線という新たな問題が生じる。ウェーハ間結線はマイクロブリッジを用いて行なう手法が提案され実装されている [3]。マイクロブリッジは、短い間隔で信頼性の高いウェーハ間結合が可能であるが、大きさが数百μmに達する。したがってウェーハ間結線数はウェーハ上のレイアウト面積に大きく影響する。超並列システムをウェーハスタック構造へ実装する場合に、ウェーハ間結線は実装上の大きな問題である。また、相互結合網は超並列システムの処理性能に対して大きな影響を及ぼす。

本研究では、ウェーハスタック構造へ階層化ネットワークを実装する場合を考え、ウェーハ間の結線が少ないネットワークである TESH(Tri connected mESHes)[4] について概説する。また、TESH や他のネットワークのウェーハ間結線数やネットワーク特性について詳しく検討する。

2 ウェーハスタック構造

2.1 ウェーハスタック

ウェーハスタック構造は、図 1に示すようにウェーハを縦に積み重ねた構造を持ち、ウェーハ間の通信はウェーハ内を貫通している結線を用いて行う。ウェーハスタック構造へネットワークをマッピングする場合には、図 2に示すようにネットワークを分割する必要がある。分割したネットワーク間のリンクは、ウェーハ間の結線となる。ウェーハスタック構造では、ウェーハ間の結線手法として、図 3のようにマイクロブリッジと呼ばれる中央を凸にした薄い金属板をウェーハの上面 (回路を構成する面) と下面に互いに直交するように取り付け、金属板の中央を接続する手法を提案している [2]。マイクロブリッジの利点は、高い信頼性でウェーハ間の間隔を短く結線できることである。その反面、図 3に示すように、マイクロブリッジに必要な幅は数百μmに達するため、数μmの幅で実現できるウェーハ上の結線と比較して非常に大きなチップ面積を必要とする。このためウェーハ間結線数の増加はチップ面積を大きく増加させ、超並列計算機システムのウェーハスタック構造への実装を困難にする。したがって、ネットワークをウェーハスタック構造にマッピングする場合には、ウェーハ間結線数をできる限り少なくする必要がある。

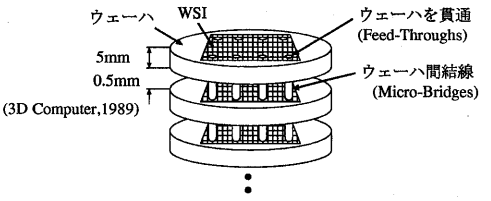

図 1: ウェーハスタック構造

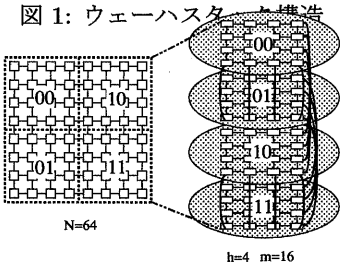

図 2: 2D メッシュのウェーハスタック構造へのマッピング

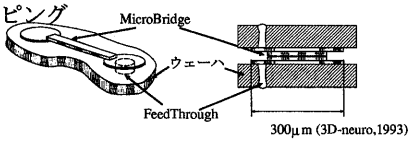

図 3: MicroBridge

2.2 ネットワーク構造表記

各ネットワークのネットワーク特性を数式で表す際に用いる表記の記号について定義を行なう。総 PE 数を N、ウェーハ枚数を h、ウェーハ 1 枚当りの PE 数を m($= \frac{N}{h}$) とする (図 2)。また、階層型ネットワークの基本モジュールの PE 数を n、レベル i を構成するレベル i-1 サブネット数を g、ウェーハ内最高階層レベルを $L_d(= 1 + \lfloor \log_g \frac{m}{n} \rfloor)$、ウェーハ 1 枚あたりの L_d の階層数を $s(= m \bmod n g^{L_d-1})$ とする。ここ

で定義した値を用いて、直径 D、ウェーハ間最大結線数 C_{max}、最大レイアウト面積 A などのネットワーク特性を表現する。

2.3 ウェーハ間最大結線数

2.3.1 2Dトーラス

2 次元トーラス (2D トーラス) は、2D メッシュの行の両端、列の両端を結合したネットワークである。2D トーラスでは、行と列がそれぞれ環結合になっている。要素数$\sqrt{N}$の環結合の直径は$\sqrt{N}/2$ なので、2D トーラスの直径は$D_{2D\text{-}torus} = \sqrt{N}$となる。

2D トーラスのウェーハ間最大結線数を定式化するため、2D トーラスを 1(PE/wafer) でウェーハスタック構造にマッピングした場合についてウェーハ間最大結線数を求める。2D トーラスは、インラインの順序で PE の配置を行なう。インラインはトーラスに用いられる配置方法で、ネットワークを中心で折り畳むように配置することで最大結線長を低く抑えることができる。$u \times u$ の 2D トーラスのインラインによる配置は以下のようになる。

$$I(N,i) = \begin{cases} \dfrac{i}{2} & i: \text{even}, \\ N - \lceil \dfrac{i}{2} \rceil & i: \text{odd} \end{cases} \quad (1)$$

このインラインを用いることによって、結線長の平均化が行なわれ、最大結線長も減少する。この配置方法を用いた場合の 2D トーラスのウェーハ間最大結線数は以下のようになる。

$$C_{max} = \begin{cases} 2\sqrt{N} + 2\sqrt{m} & \frac{N}{m} \neq 4, \\ 4\sqrt{m} & \frac{N}{m} = 4 \end{cases} \quad (2)$$

2 次元メッシュ (2D メッシュ) は、PE を格子状に並べ、上下左右 4 方向の隣合う PE 同士を結合したネットワークである。2D メッシュの直径は列と行の最大移動距離の和をとり $D_{2D\text{-}mesh} = 2(\sqrt{N} - 1)$ となる。2D メッシュでは 2D トーラスと比べ隣接 PE 間が半数のリンクで結合されている。したがって 2D メッシュのウェーハ間最大結線数は、2D トーラスの 1/2 になる。

2.3.2 ハイパーキューブ

k-次元のハイパキューブは 2^K の PE から構成され、アドレスを k bit の 2 進数で表した場合、k ビットの内 1 ビットだけ異なる PE 同士が結合される。

ハイパーキューブのルーティングは、送信元と送信先のアドレスの内、異なっているビットを下位ビットから 1 ビットずつ一致させるような経路を選択することで行なわれる。直径は、$D = k = \log_2 N$となり、平均距離は $N/2$ となる。

ハイパーキューブのウェーハスタック構造へのマッピングは以下のように行なう。まず、アドレス順に並んだネットワークを格子状に j 分割する。上位 $\log j$ビットのアドレスを参照し、この順序でウェーハに配置を行なう。この配置ではウェーハに垂直方向の PE 列を考えると、各列がそれぞれハイパーキューブを構成している。このためハイパーキューブのウェーハ間最大結線数は、(ウェーハ枚数 h に等しい PE 数を 1(PE/Wafer) で配置した場合のウェーハ間最大結線数) × (ウェーハ 1 枚当りの PE 数) となる。k 次元のハイパーキューブを m(PE/wafer) でマッピングした場合のウェーハ間最大結線数は次式で与えられる。

$$C_k = \begin{cases} \dfrac{2m(N-m)}{3} & k: \text{even}, \\ \dfrac{2N-m}{3} & k: \text{odd} \end{cases} \quad (3)$$

2.3.3 3Dトーラス

3D トーラスは 3D メッシュの対になっている面を結合したネットワークである。3D トーラスの直径は $D_{3D\text{-}torus} = (3/2)N^{\frac{1}{3}}$となる。

3D トーラスのウェーハ間最大結線数は次の 2 通りに分けることができる。1 枚以上の x-y 断面をウェーハに配置する場合は x-y 断面の PE 数に等しい。x-y 断面を分割してウェーハに配置を行なう場合は x-y 断面の PE 数と x-y 断面を分割した場合のウェーハ間最大結線数の和になる。これは $z = 1, 2, \ldots, u-1, u-2$ の x-y 断面では z 軸方向のリンク数が減少しないためである。したがって 3D トーラスのウェーハ間最

大結線数は次式で与えられる。

$$C_{max} = \begin{cases} m < N^{\frac{2}{3}} & 2(N^{\frac{1}{3}} + N^{\frac{2}{3}} + \sqrt{m}) \quad \frac{N^{\frac{2}{3}}}{m} \geq 1 \\ & 2(N^{\frac{2}{3}} + 2\sqrt{m}) \quad \frac{N^{\frac{2}{3}}}{m} = 1 \\ m \geq N^{\frac{2}{3}} & 2N^{\frac{2}{3}} \end{cases} \tag{4}$$

3Dメッシュは PE を立方格子状に並べ、隣合う PE 同士を結合したネットワークである。3D メッシュの直径は $D_{3D-mesh} = 3(N^{\frac{1}{3}} - 1)$ となる。3D メッシュの隣接 PE 間のリンク数は 3D トーラスの半数である。そのため、3D メッシュのウェーハ間最大結線数は 3D トーラスの 1/2 になる。

3 TESH

3.1 ネットワーク構造

TESH(Tori connectd mESHes)[4] は、ウェーハスタック構造を考慮した階層型ネットワークで、基本モジュール (BM) と呼ばれる 4 × 4 の 2D メッシュの間を 4×4 の 2D トーラスで結合して階層化したネットワークである (図 4)。基本モジュールをレベル 1 として、階層化を行なう毎にレベルは 1 ずつ増加するものとする。基本モジュールは、全レベルへのリンクの入出力位置があらかじめ決定されている。各レベルに対して行方向のリンク 2 本、列方向のリンク 2 本、計リンク 4 本が割り当てられている。この 4 本のリンクを用いて、基本モジュール間を 4×4 の 2D トーラスで結合し、階層化を行なうことができる。レベル 2 階層は、16 個の基本モジュールの間を 4×4 の 2D トーラスで結合して階層化を行なう。レベル 3 階層も 16 個のレベル 2 を 4×4 の 2D トーラスで結合して階層化を行なう。しかし、レベル 3 階層以降はサブネットに多数の基本モジュールを含むため、2D トーラス結合する基本モジュールの選択が必要となる。TESH では、各レベルで位置が等しい基本モジュール間を 2D トーラス結合をする。そのため、レベル 3 階層ではレベル 2 階層に含まれる BM 数 (16 個) に等しい数の 2D トーラスが並列に階層間の結合を行なう。

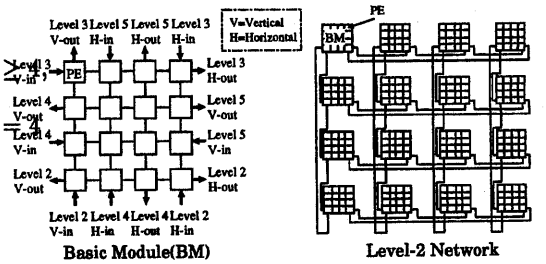

図 4: TESH のネットワーク構造

3.2 直径

TESH のルーティングは上位階層から下位階層の順序で行なわれる。最初に最上位階層のルーティングを行ない、目的のサブネットワークに到達したら、さらに下のレベルのルーティングに移る。表 1に TESH の直径を示す。

表 1: TESH の直径

L	1	2	3	4	5
N	16	256	4096	65536	1048576
D	6	19	32	43	53

3.3 ウェーハ間最大結線数

TESH はサブネットを越えて上位レベルのリンクが行なわれるため、レベル L のウェーハ間最大結線数はレベル L 以下の各レベルのウェーハ間最大結線数の和となる。TESH はサブネット内の基本モジュール数に等しい 2D トーラスで階層間の結合を行なうため、レベル L_iのみを考慮したウェーハ間最大結線数は以下のようになる。

$$\begin{aligned} C_{max}^{L_i} &= C_{max}(2Dtorus, 16, s_{L_i})16^{L_i-2} \\ &= \begin{cases} 10(16^{L_i-2}) & s_{L_i} = 1, \\ 8(16^{L_i-2}) & s_{L_i} = 4 \end{cases} \end{aligned} \tag{5}$$

ここで、レベル L_i階層のウェーハ 1 枚あたりのサブネット数を s_{L_i}で表す。したがって TESH のウェーハ間最大結線数は次式で与えられる。

$$C_{max} = C_{max}(2Dtorus, 16, s)16^{L_d-2}$$

$$+\sum_{L_i=L_d+1}^{L} 10(16^{L_i-2}) \qquad (6)$$

4 レイアウト面積

4.1 一般式

ウェーハ間最大結線数がレイアウト面積にどの様な影響を及ぼすかを調べるため、各ネットワークのレイアウト面積を求める一般式の定義を行なう。レイアウト面積とはPE、リンク、全空白領域を含めた全面積を表す。本研究ではPEの幅をW_{PE}、ウェーハ間結線(マイクロブリッジ)の幅をW_{MB}、データ線1本の幅をW_{link}、結線1本当りのデータ線数をp本、各行や列の最大トーラスリンク本数をt_x, t_yとする。この最大トーラスリンク本数とウェーハ間最大結線がレイアウト面積を求める際に用いる各ネットワークの固有の値となる。

ウェーハ間結線は熱の伝搬によってウェーハの熱分散の役割を担う。そのため、ウェーハ間結線はウェーハ上に均等に配置することが望ましい。本研究では1枚のウェーハ上に載っているネットワークを正方格子状にL分割したブロック単位でウェーハ間結線の均等配置を行なう。またウェーハ間結線の均等配置に必要な移動回数として1本のウェーハ間結線(pビット幅)に、縦横1本のリンク(pビット幅)が必要であると仮定する。また最悪の条件を仮定し、全ての均等配置に用いるリンクは同一線上に配置できないとした。つまり、均等配置に必要なリンク幅はリンク本数の和になると仮定している。ウェーハ間結線を除いたレイアウト幅W'はウェーハ間結線の均等配置のためのリンク幅を考慮して以下のようになる。

$$W' = \sqrt{m}\, W_{PE} + \sqrt{m}\, t_y\, p\, W_{link} + C_{max}\, p\, W_{link}$$

第一項はウェーハ上の全PE幅、第2項はウェーハ上のリンク幅、第3項はウェーハ間結線の分散に用いるリンク幅である。ウェーハ上のネットワークをL分割したブロック幅は$W_{BL} = \frac{W'}{\sqrt{L}}$となり、ブロック当たりのウェーハ間結線数は$c = \lceil \frac{C_{max}}{L} \rceil$となるのでブロック幅に正規化して並べたウェーハ間結線の厚さは$w = \lceil \frac{c\, p\, W_{MB}}{W_{BL}} \rceil W_{MB}$となる。したがって、ウェーハ間結線を含めたレイアウトの縦幅は$W_y = W' + w\sqrt{L}$となる。全レイアウト面積は以下のようになる。

$$A = W' \cdot W_y = W'\{W' + (\lceil \frac{c\, p\, W_{MB}}{W_{BL}} \rceil W_{MB})\sqrt{L}\}$$

図5にL=4,p=1におけるハイパーキューブのレイアウト面積Aを示す。

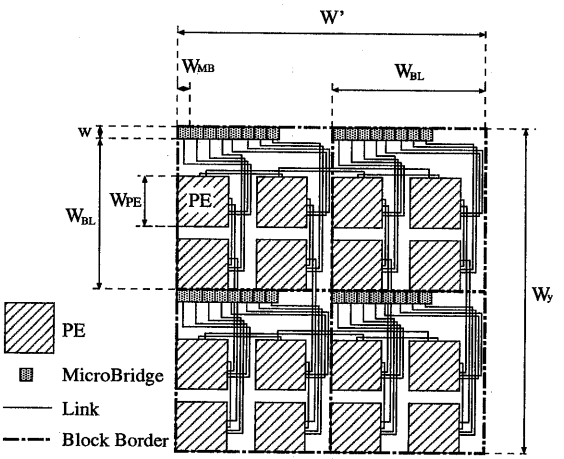

図5: ハイパーキューブのレイアウト面積

4.2 TESH

TESHの最大トーラスリンク本数は、1つの基本モジュールのみがウェーハに載っている場合、$t_x = t_y = 0$である。4つの基本モジュールがウェーハに載っている場合は、レベル2階層を構成するリンクがウェーハ内に含まれるので$t_x = t_y = 2$となる。レベル3以降ではサブネット内の基本モジュール数に等しい2次元トーラスでサブネット間の結合が行なわれるため、トーラスリンク本数は1レベル毎にサブネットの行(列)あたりの基本モジュール数だけ増加する。したがって2つ以上のサブネットが載っている最高レベルを$L_e = \lceil \log_{16} m \rceil$とすると$t_x = t_y = 2\sum_{i=Le}^{L}(4^{i-2})$となる。

5 ネットワーク特性

ウェーハスタック構造に適したネットワークの特性を検討するため、低次数の2Dトーラ

ス (2Dtorus)、3 次元構造を持つ 3D トーラス (3Dtorus)、高次数のハイパーキューブ (Hyper-Cube)、ウェーハスタック構造を考慮した階層型ネットワーク TESH の特性の比較を行なう。

ウェーハ間最大結線数は、ウェーハ枚数を 16 枚に固定して評価を行なった。面積比は、以下の条件を用いた。

●総 PE 数:4096、1 ウェーハに載せるネットワーク規模:16 × 16(= $256PEs$)、ウェーハ枚数:16 枚、PE の幅:1mm、ウェーハ上の結線 :1μm CMOS technology、マイクロブリッジの幅:500μm

ハイパーキューブは直径が小さいが、各 PE の次数が高いためウェーハ間結線数が非常に多く、ネットワーク面積も非常に大きくなる。一方、2D トーラスはウェーハ間最大結線、ネットワーク面積の増加が小さいが、ネットワーク直径の増加が大きく、PE 数が増加した場合、通信遅延が非常に大きくなる。TESH および 3D トーラスはネットワーク直径、ウェーハ間最大結線数を比較的少なくできる。特に TESH のような階層構造を有するネットワークは、少ない階層間のリンクを用いてウェーハ間結線数を減らしつつ、直径の増加を抑えることができるため、ウェーハスタック構造に最も適している。

以上の結果から、ウェーハスタック構造には低次元ネットワーク、3D トーラスのような 3 次元ネットワーク及び TESH のような階層型ネットワークが適していることが分かった。

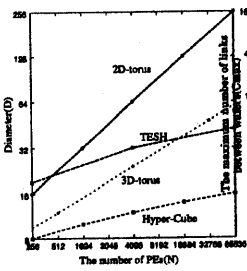

図 6: 直径

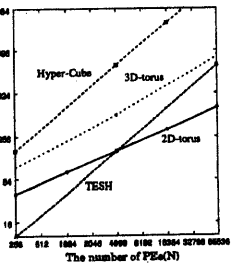

図 7: ウェーハ間最大結線数

6 まとめ

本研究では、超並列コンピュータネットワークのウェーハスタック構造インプリメンテーションについて検討した。ウェーハ間結線数を減らしつつ、直径やレイアウト面積の増加を抑えるネットワークとして、階層型ネットワーク TESH を提案した。TESH は、基本モジュール間のリンクを減らした構造になっており、従来のネットワークに比べ、ウェーハスタック構造の実装に向いたネットワークであることを示した。

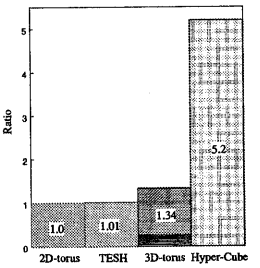

図 8: レイアウト面積 (N=4096)

今後の課題としては、階層型冗長構成法やより特性の良い 3 次元階層型ネットワークの提案やアプリケーションのマッピングによる実行性能について検討などがある。

参考文献

[1] 堀口 進:“ウェハ規模超密度集積回路について”,Hybrids,6,1,pp.16-21(1990).

[2] Little.M.J,J.Grinberg:
“The 3-D Computer:An Integrated Satck of WSI Wafers,in Wafer Scale Integration”,Kluwer Academic Publishers,pp.253-317,1989.

[3] M.L.Campbell,S.T.Toborg:“3D wafer stack neurocomputing”,Proceedings of International Conference on Wafer Scale Integration,pp.67-74,1993.

[4] V.K.Jain,T.Ghirmai,S.Horiguchi:“Reconfiguration and field for TESH:A New Hierarchical Interconnection Network for 3-D Integration”,Proc. IEEE Int'l Conf. SISI(1996).