

修士学位論文

論文題目 ハイパキューブ結合型マルチ
プロセッサシステムのWSI構成法
に関する研究

提出者 東北大学大学院工学研究科
情報工学 専攻

学籍番号 3m271

氏名 福田大

修士学位論文

ハイパキューブ結合型マルチプロセッサ
システムのWSI構成法に関する研究

東北大学大学院 工学研究科
情報工学専攻

福田 大

目次

1	序論	1
1.1	研究の背景と目的	1
1.2	WS I の欠陥救済技術	2
1.3	本論文の構成	2
2	超並列向きのプロセッサ間接続方式	5
2.1	はじめに	5
2.2	直接網のネットワーク	5
2.3	ハイパキューブ	6
2.4	ハイパキューブ結合型ネットワーク	7
2.4.1	巡回型ハイパキューブ	7
2.4.2	ハイパネット	9
2.5	新しく提案するハイパキューブ結合型トポロジ	11
2.6	特性の比較	12
2.7	まとめ	13
3	ハイパキューブ結合型マルチプロセッサシステムのWS I 構成法	24
3.1	はじめに	24
3.2	本研究で提案するWS I 構成法	25
3.2.1	ハイパキューブの構成法	26
3.2.2	巡回型ハイパキューブの構成法	27
3.2.3	ハイパネット	29
3.2.4	階層型ハイパキューブ	29
3.3	まとめ	31
4	WS I システムの性能評価	42
4.1	はじめに	42
4.2	レイアウト面積	42

4.3	歩留まり	44
4.4	まとめ	47
5	結論	56
5.1	本研究のまとめ	56
5.2	今後の課題	57

目 次

2.1	完全結合網とクロスバ網	15
2.2	代表的な直接網のトポロジ	16
2.3	HC(2,4)	17
2.4	CCC(4,3)	18
2.5	HN(3,2)	19
2.6	HHC(2,2,3)	20
2.7	直径の比較	21
2.8	総リンク数の比較	22
2.9	平均距離の比較	23
3.1	階層型冗長構成法	32
3.2	HC(2, d) のPE配置	33
3.3	HC(2,4) のWSI構成例	34
3.4	HC結合リンク束	35
3.5	スイッチの切り換えパターン	35
3.6	CCC(c, d) のPE配置	36
3.7	CCC(4,2) のWSI構成例	37
3.8	HN(d, h) のPE配置	38
3.9	HN(2,3) のWSI構成例	39
3.10	HHC(d_1, d_2, h) のPE配置	40
3.11	HHC(2,2,2) のWSI構成例	41
4.1	通信制御回路を考慮した場合のレイアウト面積の比較	51
4.2	HCのレイアウト面積に対するレイアウト面積の比較	52
4.3	通信制御回路を考慮しない場合のレイアウト面積の比較	53
4.4	通信制御回路を考慮した場合としない場合のレイアウト面積の差	54
4.5	通信制御回路を考慮した場合の歩留まりの比較	55

表 目 次

1.1	欠陥救済技術の分類	4
2.1	代表的直接網の直径	14
2.2	ネットワークの特性の一覧	14
4.1	各ネットワークのPE数とリンク数	49
4.2	各構成法のPE配列	50
4.3	各ネットワークのレイアウト面積	50

第 1 章

序論

1.1 研究の背景と目的

近年、コンピュータの処理能力は飛躍的に向上してきたが、逐次処理方式に基づく従来のノイマン型コンピュータの性能の向上も物理的な限界に近づきつつある。しかし、科学技術計算をはじめとして、コンピュータに対する処理能力の要求はますます増加し、新しいアーキテクチャによるコンピュータの実現が求められている。処理を大幅に高速化させる最も自然な考え方は処理の並列化であり、(超)並列システムが注目されている。並列システムではプロセッサ間の通信が必要となるが、プロセッサ数が 10^6 個を超えるような超並列システムでは膨大な数のプロセッサ間での通信が必要となる。そのため、従来の多数のチップを用いた構成法ではチップ間での配線、通信遅延、及び、信頼性が問題となってくるばかりではなく、その大きさ自体も問題になってくる。そこで、VLSI チップの集積規模を上げ、1つのチップに複数のプロセッサ等からなる回路を搭載させることにより、チップ間の配線を減らし、システムを小型化させることができる。

ウェハスケールインテグレーション (Wafer Scale Integration: 以下WSI と呼ぶ) は、このチップ面積をウェハ規模にまで拡大して1枚のウェハ上に大規模なシステム全体を搭載させようとするものである。1枚のウェハ上に全システムを搭載させることにより、システムの小型化、高速化、高信頼性化等が期待できる [1] [2] [3]。

小型で高速な超並列システムの実現には、このウェハスケール集積システムが不可欠であると考えられる。

現在の集積回路製造技術では、回路の製造の各段階でチリや微生物などにより、多少の欠陥 (defect) が発生することは避けられない。この欠陥はチップ面積の拡大とともに増えるので、良品チップが得られる確率はチップ面積の拡大とともに急激に低下する。従って、1枚のウェハ上に大規模なシステムを搭載するWSI では良品はほとんど得られないことになる。そこで、WSI の実現の最大の問題点は、この欠陥にどのように対処するかとい

うことになる。WSIシステムでは、膨大なプロセッサや配線領域の中で欠陥が発生するため、プロセッサやリンク、スイッチなどの回路の一部分に冗長性を持たせて、システムの製造の最終段階、あるいは使用中のある時期に欠陥を回避して(ディフェクト・トレランス: defect tolerance) システムを再構成させる。このためレイアウトもそのような再構成が可能ないように設計する必要がある。しかし、ウェハ上で構成させるネットワーク構造が大規模かつ複雑になってきており、新しい構成法の研究の重要性が高まっている。

(超)並列システムのWSI構成に関する研究は、プロセッサを2次元格子状に並べた構造に関するものが多く、様々な再構成アルゴリズムが提案されている [4][5]。しかし、最近ではもっと複雑な構造のWSI構成法に関する研究も行われており、ハイパキューブ結合に関するものは伊藤らにより検討されている [6][7][8]。

ハイパキューブ結合は汎用の並列システムのプロセッサ間接続方式として注目されているが、超並列システムを目指す場合、1つのプロセッサ当たりのリンク接続数と総リンク数が増加し、物理的に実現が困難となる。そこで、ハイパキューブの1つのプロセッサ当たりのリンク接続数と総リンク数を減らした構造がいくつか提案されている [9][10]。

本研究の目的は超並列を目指したハイパキューブ結合型のマルチプロセッサシステムのWSI構成法を検討することである。また、ハイパキューブ結合型のマルチプロセッサシステムとして、ハイパキューブを階層的に組み合わせたネットワーク構造を新たに考案し、ハイパキューブや他のハイパキューブ結合型のマルチプロセッサシステムとの比較も行った。

1.2 WSIの欠陥救済技術

WSIにおける欠陥救済技術は冗長回路構成法、欠陥の検出法、および、実際の配線の切り換え手法の3つに分けられる [11]。冗長回路構成法は回路上に欠陥が発生することを見こして回路に冗長回路を付加する構成法である。欠陥個所は冗長回路で直接置き換えたり、回路を階層的にして下位の階層の欠陥を上位の階層で救済させたりする。また、実際に欠陥を救済してシステムを再構成させるアルゴリズムも必要である。ウェハ上に搭載させる回路が大規模であるので、外部からテストなどで欠陥を検出することは困難であり、回路の内部に自己テスト回路を搭載させることが必要となる。実際の欠陥部分の救済は配線の切り換えにより行い、スイッチやヒューズを用いて電気的に行う方法や、レーザー光を用いて光学的に行う方法がある。表 1.1にWSIの実用化のための欠陥救済技術の分類を示す。

本研究は回路に冗長回路を付加させる冗長回路構成法に関する研究である。

1.3 本論文の構成

本論文の構成は以下の通りである。

第2章では超並列システム向きのプロセッサ間接続方式であるハイパキューブ結合型のネットワークについて述べる。また、新しく提案するハイパキューブを階層的に組み合わせた構造について述べ、他のネットワークと特性の比較を行う。第3章では本研究で提案するWSI構成法について述べる。第4章ではWSI構成法の評価法について述べ、評価を行う。第5章は結論であり、また今後の課題について述べる。

表 1.1: 欠陥救済技術の分類

冗長回路構成	冗長構成法	冗長ブロック置換 階層的冗長構成法
	再構成アルゴリズム	
欠陥の検出	自己テスト	テスト回路の内蔵
配線の切り換え	電気的手法	バイパス回路 切り替え回路 ヒューズ
	光学的手法	レーザーカット レーザーコネクト

第 2 章

超並列向きのプロセッサ間接続方式

2.1 はじめに

並列システムのプロセッサ間接続方式(相互結合網: interconnection network)は、そのトポロジ(topology: 幾何学的形状)により直接網(direct network)と間接網(indirect network)に分けられる。前者はプロセッサ間の接続が静的に固定されているもので、静的網(static network)とも呼ばれる。また、後者はプロセッサ間の接続をスイッチの制御により動的に行うもので、動的網(dynamic network)とも呼ばれる。直接網および間接網は、それぞれ図 2.1に示すような完全結合網(completely connected network)およびクロスバ網(cross-bar network)を規則的に簡略化したものと見ることができる。

本研究では直接網の中で超並列向きのプロセッサ間接続方式として注目されているハイパキューブと、ハイパキューブを他の直接網のトポロジと組み合わせたハイパキューブ結合型のプロセッサ間接続方式について考える。

本章では2.2節で直接網の直接網のマルチプロセッサシステムについて述べる。また、2.3節でハイパキューブについて述べる。2.4節では、ハイパキューブ結合型のプロセッサ間接続方式について述べる。さらに、2.5節でハイパキューブを階層化させて次数とリンク数を減少させた、新しいネットワークを提案する。2.6節ではこれらのハイパキューブ結合型ネットワークの特性の比較を行う。

以下では直接網のトポロジによるマルチプロセッサシステムを単にネットワークと呼ぶ。

2.2 直接網のネットワーク

図 2.2に代表的な直接網のトポロジの例を示す。図で黒丸はプロセッサ要素(processing element: 以下PEと記す)を表し、PE間を結ぶ線はPE間の双方向通信を行うためのリンクを表している。

直接網のネットワークにおいて、2つのPE間の通信を行う時に経由するリンク数、あるいは経由するPE数に1を加えた値をそれらのPE間の距離と呼ぶ。また、距離の最大値をそのネットワークの直径(diameter)と呼ぶ。超並列システムでは、多数のプロセッサ間でデータのやりとり等の通信を行う必要があり、この直径を小さくすることが重要になってくる。また、通信経路を集中させないことも必要である。

表2.1に代表的な直接網のネットワークのPEの総数 N に対する直径の比較を示す。ハイパキューブはこの直径が $\log_2 N$ となり、これらのネットワークの中で最小となる。また、ハイパキューブはバイナリトリーのように通信経路が集中することはない。したがって、ハイパキューブは超並列向きのネットワーク構造と言える。

本研究では、直接網のネットワークの特性を示すものとして、直径を含めて以下に挙げるものを考える。

- 次数
- 直径
- 平均距離
- 総リンク数
- ルーティング法

次数は、1つのPE当たりのリンク接続数を表す。平均距離はPE間距離の平均を表す。本研究では、任意の2つのPE間の通信の確率が全て等しい場合のみを考える。総リンク数は通信リンクの総数を表す。ルーティング(routing)法は通信元のPEから通信先のPEへの経路(パス)を決めるアルゴリズムのことである。

2.3 ハイパキューブ

ハイパキューブ(hypercube: 以下HCと記す)は、 n -キューブ、あるいは超立方体とも呼ばれ、汎用の(超)並列計算機の結合方法として注目されており、実際にいくつかの並列システムに採用されている[12][13]。通常HCといえは2進HC、つまりバイナリキューブを指し、本論文でもこのバイナリキューブを意味している。

d 次元のハイパキューブを $HC(2,d)$ で表すことにする。 $HC(2,d)$ は 2^d 個のPEからなり、各PEのアドレスを d ビットの2進数で表すと、 d ビットのアドレスのうちの1ビットだけが異なるPE同士が結合される。従って、 $HC(2,d)$ の次数は d 、総リンク数は $d \cdot 2^{d-1}$ 本となる。

図2.3に $HC(2,4)$ を示す。

HCのルーティング法は簡単である。HC(2, d)の場合、例えば通信元のPEと通信先のPEの d ビットのアドレスのうち、一致していないビットを下位のビットから1ビットずつ一致させるように経路を決めればよい。具体的には、図2.3のHC(2,4)で(0000)のアドレスを持つPEから(1111)のアドレスを持つPEへの通信経路は(0000) → (0001) → (0011) → (0111) → (1111)となる。

このルーティング法により、HC(2, d)の直径、平均距離は以下に示すとおりとなる。

$$\text{直径} = d \quad (2.1)$$

$$\text{平均距離} = \frac{\sum_{i=0}^d \binom{d}{i} i}{2^d} = \frac{d}{2} \quad (2.2)$$

2.4 ハイパキューブ結合型ネットワーク

ハイパキューブはメッシュやトリーなどのトポロジに比べて、PE数に対する直径を小さくできる。しかし、PE数が 10^6 個を超える超並列システムを目指す場合、それでもPE間の直径は小さいとはいえず、PE間の通信遅延によりシステム全体の性能が制限されてくる。また、PE当たりのリンク接続数が増えて総リンク数が膨大となり、実現が困難になる。

一般的に、次数を増やせば直径は小さくなり、次数を減らせば直径が長くなる。次章でも述べるように、超並列を目指す場合総リンク数が問題となってくるので、本研究ではHCの次数を減らして総リンク数を減らしたネットワーク構造を検討する。

HCと他のトポロジを階層的に組み合わせ、次数と総リンク数を減らしたネットワーク構造がこれまでにいくつか提案されている。以下にそのうちの代表的なネットワーク構造について説明する。

2.4.1 巡回型ハイパキューブ

巡回型ハイパキューブ (cube-connected cycles: 以下CCCと記す) はHCの各PEをリングで置き換えた構造であり、1981年に Preparata と Vuillemin によって提案された [9]。

c 個のPEを1つのサイクルとして、このサイクルを d 次元のハイパキューブ結合させた構造を $c \cdot 2^d$ のCCCと呼び、を $CCC(c,d)$ と表すことにする($c \geq d$)。CCC(c,d)の次数は3、PE数は $c \cdot 2^d$ 、総リンク数は $c \cdot 2^d + d \cdot 2^{d-1}$ となる。

CCCの各PEに対してアドレス(l,p)を与える。ここで l はサイクルのアドレスを表し、 c ビットの2進数で表される。また、 p はサイクル内のPEのアドレスを表し、0から $d-1$ までの10進数で表される。

各PEは3つの入力ポート F (Forward), B (Backward), L (Lateral) を持ち、 $0 \leq p \leq d-1$ のアドレスを持つPEは F, B の2つのポートがサイクル内での結合に使われ、 L ポートはサイクル間の結合に使われる。また、 $d \leq p \leq c-1$ のアドレスを持つPEは F, B のポートにより、サイクル内の結合のみが行なわれる。PE間の結合はそのアドレス (l, p) により次の関係で決定される。

$$\left. \begin{aligned} F(l, p) &= B(l, (p+1) \bmod h) \\ B(l, p) &= F(l, (p-1) \bmod h) \\ L(l, p) &= L(l + \epsilon \cdot 2^p, p) \end{aligned} \right\} \quad (2.3)$$

ここで、 $\epsilon = 1 - 2\text{bit}_p(l)$ であり、 $\text{bit}_p(l)$ は l の下位から p 番目のビットの数である。

図2.4に $CCC(4,3)$ を示す。図でサイクル内のアドレスが3であるPEは、サイクル内の結合のみが行われている。

CCCのルーティング法は[14]に示されており、簡単にまとめると次のようになる。ここで、通信元のアドレスを (l_s, p_s) 、通信先のアドレスを (l_o, p_o) とする。移動中のアドレスを (l_i, p_i) とし、 i は s から始める。

- (1) l_i が l_o と一致していれば、 F ポートあるいは B ポートの結合によりサイクル内の移動を行い終了する。
- (2) (l_i, p_i) に L ポートの結合が行われていなければ、 F ポートの結合により、 L ポートの結合が行われているPEまで移動する。そして $\text{bit}_p(l_i)$ が $\text{bit}_p(l_o)$ と一致していなければ、 L ポートの結合によりサイクルを移動し、(1)に戻る。
- (3) $\text{bit}_p(l_i)$ が $\text{bit}_p(l_o)$ と一致していれば、 F ポートの結合によりサイクル内で1回移動し、(2)に戻る。

具体的には、図2.4の $CCC(4,3)$ で $(000,0)$ のアドレスのPEから $(111,3)$ のアドレスのPEへの通信経路は、 $(000,0) \rightarrow (001,0) \rightarrow (001,1) \rightarrow (011,1) \rightarrow (011,2) \rightarrow (111,2) \rightarrow (111,3)$ となる。

また上のルーティングアルゴリズムによるCCCの直径、平均距離の計算も[14]で行われている。

$$\text{直径} = d + (d-1) + (c-d) + \frac{c}{2} = d-1 + \frac{3}{2} \quad (2.4)$$

$$\begin{aligned} \text{平均距離} &= \frac{d}{2} + (d-1) + (c-d) - \sum_{k=1}^{d-1} \frac{k}{2^k} + \frac{c}{4} \\ &= \frac{d}{2} + (d-1) + (c-d) - \frac{2^d - (d+1)}{2^{d-1}} + \frac{c}{4} \\ &= \frac{d}{2} - 3 + \frac{d+1}{2^{d-1}} + \frac{5}{4}c \end{aligned} \quad (2.5)$$

上式において、直径の計算式の最初の d はサイクル間の移動で使用するハイパキューブ結合のリンクの最大数を表している。次の $(d-1) + (c-d)$ はサイクル内での移動で使用するリンクの最大数を表し、 $c/2$ は通信先のサイクル内での移動で使用するリンクの最大数を表している。また、平均距離の計算式の最初の $d/2$ はサイクル間の移動で使用するハイパキューブ結合のリンク数の平均を表している。次の $(d-1) + (c-d) - \sum_{k=1}^{d-1} (k/2^k)$ はサイクル内の移動で使用するリンク数の平均を表している。ここで \sum の項は通信元と通信先のサイクルアドレスの一致により、ハイパキューブ結合のリンクを使用しなくてもよいリンク数の期待値を表す。厳密には L ポートの結合のない PE を経由しなくてもよい場合も生じるが、この場合は少ないので、ここでは考えないことにしている。また、 $c/4$ は通信先のサイクル内での移動で使用するリンク数の平均を表している。

2.4.2 ハイパネット

ハイパネット (hypernet: 以下 HN と記す) はハイパキューブ、トリー、バス結合を基本ネット (クラスタ: cluster) として、それらを階層的に完全結合してできる構造であり、1987 年に Hwang と Ghosh によって提案された [10]。ここではハイパキューブをクラスタとしたハイパネットについてのみ考える。

d 次元の HC をクラスタとした h 階層の HN を $HN(d, h)$ で表すことにする。 $(d, 1)$ はクラスタを表す。 $HN(d, h)$ の各 PE は d 本のクラスタ内のハイパキューブ結合用リンクと 1 本のクラスタ間の完全結合用のリンクを持つ。従って、 $HN(d, h)$ 次数は $d+1$ となる。

$HN(d, h)$ のアドレスを N_h ビットの 2 進数で表す。 $HN(d, h)$ は次の規則により複数の $HN(d, h-1)$ から構成される。

- (1) N_h の下位の $(h-1)$ ビットが $011 \dots 1$ のビット列となる PE (例えば、 $HN(d, 2)$ の場合下位 1 ビットのアドレスが 0 で、 $HN(d, 3)$ の場合下位 1 ビットのアドレスが 01 の PE) が $HN(d, h-1)$ 間の完全結合に使用される。
- (2) $HN(d, h)$ のアドレスは $2\{N_{h-1} - (h-1)\} + (h-1)$ ビットとなり、上位の $N_{h-1} - (h-1)$ ビットを次の $N_{h-1} - (h-1)$ ビットと入れ換えたアドレスを持つ PE 同士が結合される。また、上位の $N_{h-1} - (h-1)$ ビットと次の $N_{h-1} - (h-1)$ ビットが同じ場合は結合を行わない。したがって、 $HN(d, h)$ を構成する $HN(d, h-1)$ (サブネット、あるいは $(d, h-1)$ ネットという) の数は、 $2^{N_{h-1} - (h-1)}$ 個となる。

上の規則より、

$$\left. \begin{aligned} N_h &= 2\{N_{h-1} - (h-1)\} + (h-1) = 2N_{h-1} - (h-1) \\ N_1 &= d \end{aligned} \right\} \quad (2.6)$$

であり、この式より、 $N_h = 2^{h-1} \cdot (d-2) + h + 1$ となる。このアドレスの上位から $2^{k-2} \cdot (d-2) + 1$ ビットずつのアドレスは順に、 $(d, k-1)$ ネットのアドレスを表している ($k = h, h-1, \dots, 2$)。また、下位の d ビットはクラスタ内のアドレスを表している。ここで、 $\sum_{k=2}^h \{2^{k-2} \cdot (d-2) + 1\} + d = 2^{h-1} \cdot (d-2) + h + 1$ となっている。従って、 $HN(d, h)$ のPE数は $2^{2^{h-1} \cdot (d-2) + h + 1}$ 、サブネット数は $2^{N_{h-1} - (h-1)} = 2^{2^{h-2} \cdot (d-2) + 1}$ となる。

また、 $HN(d, h)$ の総リンク数は、 $HN(d, h)$ の総リンク数およびサブネット数をそれぞれ、 L_h および S_h とすると以下の式より決まる。

$$\left. \begin{aligned} L_h &= L_{h-1} \cdot S_h + \frac{S_h \cdot (S_h - 1)}{2} \quad (h \geq 2) \\ L_1 &= d \cdot 2^{d-1} \\ S_h &= 2^{2^{h-2} \cdot (d-2) + 1} \quad (h \geq 2) \end{aligned} \right\} \quad (2.7)$$

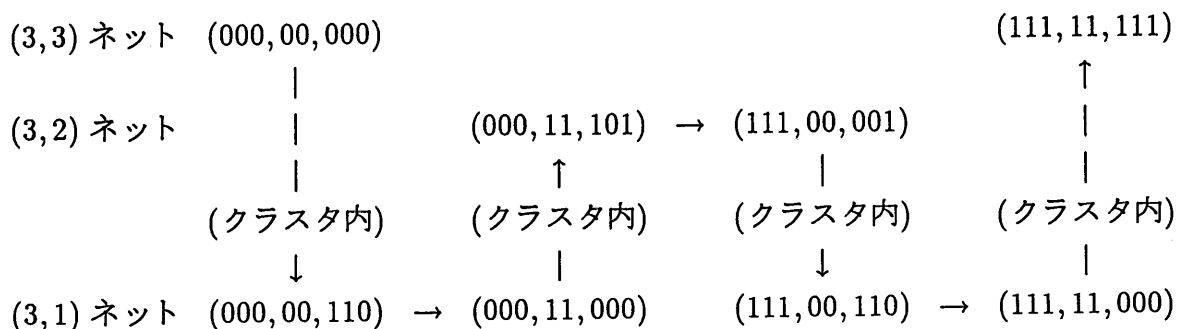
図2.5に $HN(3,2)$ を示す。図で $HN(3,1)$ で示した点線の中がクラスタであり、 $HN(3,2)$ は4つの $HN(3,1)$ をサブネットとして構成される。またサブネット間の完全結合はそれぞれのアドレスの下位の1ビットが0のアドレスのPE間で行われている。

HN のルーティング法も[10]で述べられており、以下のようにまとめられる。ここで、 $HN(d, h)$ において通信元のアドレスを n_s 、通信先のアドレスを n_o とし、移動中のアドレスを n_i とする。 n_s, n_o, n_i はそれぞれ、 $2^{h-1} \cdot (d-2) + h + 1$ ビットの2進数である。

- (1) n_i と n_o の $(d, k-1)$ ネットのアドレスを上位から順に比較する ($k = h, h-1, \dots, 2$)。
 $(d, k-1)$ のアドレスが異なっていれば(3)を実行する。
- (2) クラスタ内での移動を行い終了する。
- (3) $(d, k-1)$ ネット間の完全結合リンクを利用して $(d, k-1)$ ネットのアドレスが同じとなるように移動する。このとき、完全結合リンクを利用するために、さらに $(d, k-2)$ ネット以下のネット間で移動する必要がある。

具体的には、図2.5の $HN(3,2)$ で(00,000)のアドレスのPEから、(11,111)のアドレスのPEへの通信経路は、(00,000) → (00,010) → (00,110) → (11,000) → (11,001) → (11,011) → (11,111)となる。

また $HN(3,3)$ で、(000,00,000)のアドレスのPEから(111,11,111)のアドレスのPEへの通信経路は以下のようにして求められる。



このルーティング法による、HN(d, h)の直径、および平均距離をそれぞれ D_h, \bar{d}_h とする。 h 階層のHNの直径、および平均距離は、それぞれ2つのサブネットの直径、あるいは平均距離と2つのサブネット間の完全結合リンクの和となり、以下の関係が成り立つ。

$$\left. \begin{aligned} D_h &= 2D_{h-1} + 1 \\ D_1 &= d \end{aligned} \right\} (2.8) \quad \left. \begin{aligned} \bar{d}_h &= 2\bar{d}_{h-1} + 1 \\ \bar{d}_1 &= d/2 \end{aligned} \right\} (2.9)$$

これらから、

$$\text{直径 } (D_h) = 2^{h-1} \cdot (d+1) - 1 \quad (2.10)$$

$$\text{平均距離 } (\bar{d}_h) = 2^{h-2} \cdot (d+2) - 1 \quad (2.11)$$

となる。

2.5 新しく提案するハイパキューブ結合型トポロジ

d_1 次元のHCをクラスタとして、このクラスタを h 階層に d_2 次元のHC結合させた構造を階層型ハイパキューブ (hierarchical hypercube: 以下HHCと記す) と呼び、HHC(d_1, d_2, h)で表すことにする ($h \leq 2^{d_1} + 1$)。

HHC(d_1, d_2, h)は 2^{d_2} 個のHHC($d_1, d_2, h-1$)からなる。また、HHC($d_1, d_2, 1$)はクラスタであり、 d_1 次元のハイパキューブである。 $h=2$ (つまり、 d_1 次元のHCの代表PEを d_2 次元のHC結合させた場合)の構造は[15]で検討されている。

各PEはクラスタ内の結合と1つの階層のクラスタ間の結合を行う。クラスタ内のPE数は 2^{d_1} であり、 2^{d_1} 階層までの階層化を行うことができる。また、各PEには $(h-1) \cdot d_2 + d_1$ ビットの2進アドレスを与える。上位の $(h-1) \cdot d_2$ ビットはクラスタのアドレスを表し、下位の d_1 ビットはクラスタ内のアドレスを表す。下位から i ビットめの数を b_{i-1} で表すものとする。 $b_{(h-1) \cdot d_2 + d_1 - 1}, \dots, b_{(h-2) \cdot d_2 + d_1}$ の d_2 ビットは h 階層の結合を表すアドレスであり、 h 階層の結合はクラスタ内のアドレスが10進数で $h-2$ であるPEに対して行われる。

従って、HHC(d_1, d_2, h)のPE数は $2^{d_1 + (h-1) \cdot d_2}$ 、次数は $d_1 + d_2$ となる。また、HHC(d_1, d_2, h)のクラスタ数は $2^{(h-1) \cdot d_2}$ であるので、クラスタ内のリンク数は $d_1 \cdot 2^{d_1-1} \cdot 2^{(h-1) \cdot d_2}$ 、クラスタ間のリンク数は $(h-1) \cdot d_2 \cdot 2^{(h-1) \cdot d_2 - 1}$ となり、総リンク数は $d_1 \cdot 2^{d_1 + (h-1) \cdot d_2 - 1} + (h-1) \cdot d_2 \cdot 2^{(h-1) \cdot d_2 - 1}$ となる。

図2.6にHHC(2,2,3)を示す。図で、HHC(2,2,1)で示した点線の丸の中がクラスタであり、 2^2 個のクラスタ内の(00)のアドレスを持つPE同士を、2次元のハイパキューブ結合させて、HHC(2,2,2)を構成させる。HHC(2,2,3)は 2^2 個のHHC(2,2,2)内の各クラスタの(01)のアドレスを持つPEをそれぞれ2次元のハイパキューブ結合をさせて構成させる。

HHCのルーティング法は次のようになる。HHC(d_1, d_2, h)において、通信元のアドレスを n_s 、通信先のアドレスを n_o とし、移動中のアドレスを n_i とする。 n_s, n_o, n_i はそれぞれ、 $(h-1) \cdot d_2 + d_1$ ビットの2進数である。

- (1) n_i と n_o の h 階層のアドレスを上位から順に d_2 ビットごとに比較する。 h 階層のアドレスが異なっていれば(つまり、 $b_{(h-1) \cdot d_2 + d_1 - 1}, \dots, b_{(h-2) \cdot d_2 + d_1}$ のアドレスが異なっている)、クラスタ内のアドレスが $h-2$ であるPEへ移動し、そこからクラスタ間のハイパキューブリックを利用して同じアドレスを持つ階層に移動する。
- (2) クラスタ内での移動を行い終了する。

具体的には、図2.6のにHHC(2,2,3)で(00,00,00)のアドレスのPEから(11,11,11)のアドレスのPEへの通信経路は、(00,00,00) → (00,00,01) → (01,00,01) → (11,00,01) → (11,00,00) → (11,01,00) → (11,11,00) → (11,11,01) → (11,11,11)となる。

このルーティング法による直径および平均距離は次のようになる。

$$\text{直径} = (h-1) \cdot (d_1 + d_2) + d_1 \quad (2.12)$$

$$\text{平均距離} = (h-1) \cdot \left(\frac{d_1}{2} + \frac{d_2}{2} \right) + \frac{d_1}{2} \quad (2.13)$$

直径の $(h-1) \cdot (d_1 + d_2) + d_1$ は1つの階層ごとのクラスタ内とクラスタ間の最大リンク数を表している。最後の d_1 は同じクラスタまで移動してからの最大リンク数を表す。また平均距離の $(h-1) \cdot (d_1/2 + d_2/2)$ は1つの階層ごとのクラスタ内とクラスタ間のリンク数の平均を表し、最後の $d_1/2$ は同じクラスタまで移動してからのリンク数の平均を表す。

2.6 特性の比較

表2.2に2.2節で挙げたネットワークの特性のうちの次数、直径、平均距離、および総リンク数を、HC(2, d)、CCC(c , d)、HN(d , h)、およびHHC(d_1 , d_2 , h)についてまとめたものを示す。

CCCの次数は c および d の値によらず3と一定であり、これらのネットワークの中で最も小さくなる。また、HNとHHCの次数は階層数を増やすとHCよりも小さくすることができる。

PE数に対する各ネットワークの直径、総リンク数、および平均距離の変化をグラフにしたものを、それぞれ図2.7、図2.8、および図2.9に示す。それぞれの値はCCCでは、(c , d)を(1,1), (2,2), (3,3), ...と変化させていったときの値を示している。また、HNとHHCでは5次元のHCをクラスタとして、階層数を増やしていったときの値を示している。

直径についてはHCが最小であり、次にHN、HHC、CCCの順になっている。また、CCCはHCの約2倍となっている。これはCCCの次数が小さくなっているために、最も遠いPE間を通信するときの中継するPEの数がHCの場合の約2倍となるということである。

次に総リンク数については、HCが最大、CCCが最小、HNとHHCは同程度となっている。また、平均距離については直径と異なり、小さい順にHC、HHC、HN、CCCの順になっている。HHCはHNよりも大きい直径を持つが、平均距離は逆に小さくなっている。

以上を各ネットワークについてまとめると次のようになる。

- HCは直径、平均距離は最も小さいが、次数と総リンク数は最も大きくなる。
- CCCは次数と総リンク数は最小となるが、直径と平均距離はHCに比べてかなり大きくなる。
- HNはHCとCCCの中間的な次数、直径、総リンク数、および平均距離を持つ。階層数を増やすとPE数が急激に増加し、ネットワークのサイズの自由度が小さい。
- HHCはHCとCCCの中間的な次数、直径、総リンク数、および平均距離を持つが、平均距離がHNよりも小さくなる。

また、本研究では各ネットワークのルーティング法については特に比較は行わなかったが、HNのルーティングは再帰的であり、良いアルゴリズムとはいえない。

2.7 まとめ

本章では超並列システム向きのプロセッサ間接続方式として注目されているハイパキューブ、およびハイパキューブ結合型のネットワーク構造について述べた。また、新しいハイパキューブ結合型のネットワークとして、階層型ハイパキューブ(HHC)を提案して、他のハイパキューブ結合型のネットワークと比較を行った。その結果、階層型ハイパキューブはハイパネットとほぼ同程度のネットワーク特性を持つが、クラスタ内とクラスタ外のハイパキューブ結合の次数を変化させることにより様々なサイズのネットワークを作れるという点と、ルーティングが直接的であるという点により、ハイパネットよりも優れているといえる。

また、階層型ハイパキューブでの並列アルゴリズムの展開法は今後の課題である。

表 2.1: 代表的直接網の直径

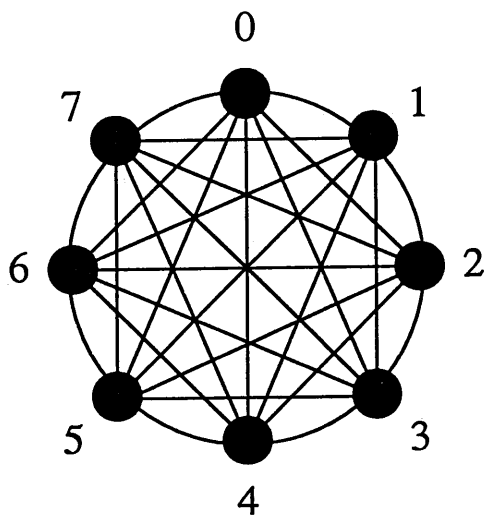
トポロジ	直径
リング	$N/2$
2次元メッシュ	$2\sqrt{N} - 2$
2次元トーラス	\sqrt{N} あるいは $\sqrt{N} - 1$
バイナリトリー	$2\log_2 N$
バイナリハイパキューブ	$\log_2 N$

表 2.2: ネットワークの特性の一覧

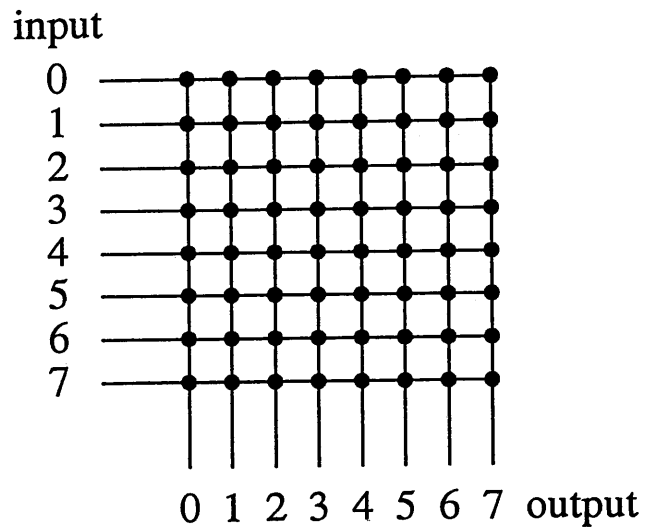
	HC(2,d)	CCC(c,d)
PE数	2^d	$c \cdot 2^d$
次数	d	3
直径	d	$d - 1 + 3c/2$
総リンク数	$d \cdot 2^{d-1}$	$c \cdot 2^d + d \cdot 2^{d-1}$
平均距離	$d/2$	$d/2 - 3 + (d+1)/2^{d-1} + 5c/4$

	HN(d,h)	HHC(d ₁ ,d ₂ ,h)
PE数	$2^{2^{h-1} \cdot (d-2) + h + 1}$	$2^{d_1 + (h-1) \cdot d_2}$
次数	$d + 1$	$d_1 + d_2$
直径	$2^{h-1} \cdot (d + 1) - 1$	$(h - 1) \cdot (d_1 + d_2) + d_1$
総リンク数	注	$d_1 \cdot 2^{(h-1) \cdot d_2 + d_1 - 1} + (h - 1) \cdot d_2 \cdot 2^{(h-1) \cdot d_2 - 1}$
平均距離	$2^{h-2} \cdot (d + 2) - 1$	$(h - 1) \cdot (d_1/2 + d_2/2) + d_1/2$

注)(2.7)より求まる。

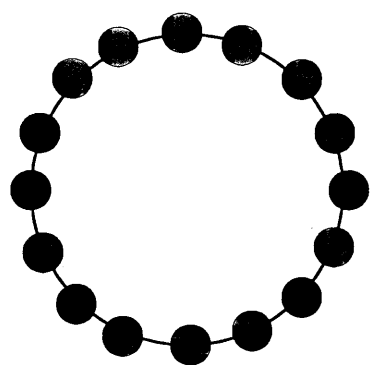


completely connected network

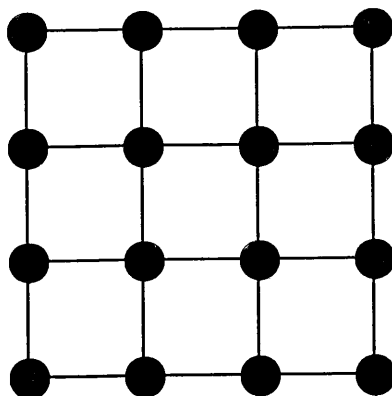


cross-bar network

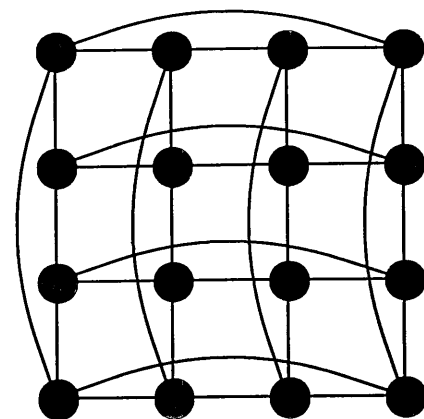
図 2.1: 完全結合網とクロスバ網



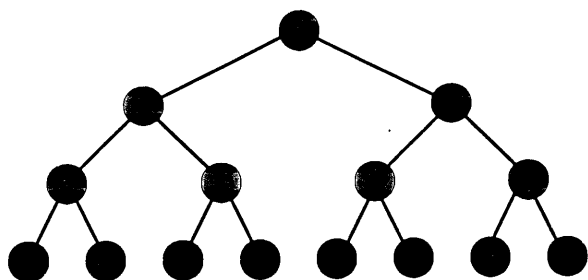
Ring



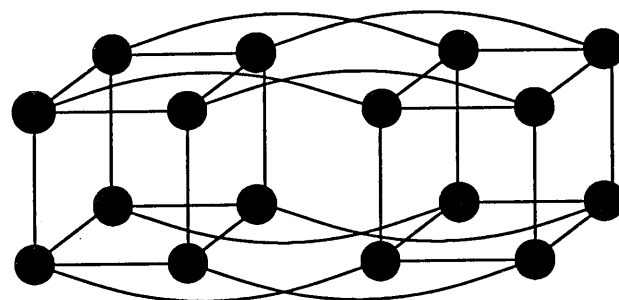
2-dimensional-Mesh



2-dimensional-Torus



binary-Tree



binary-Hypercube

図 2.2: 代表的な直接網のトポロジ

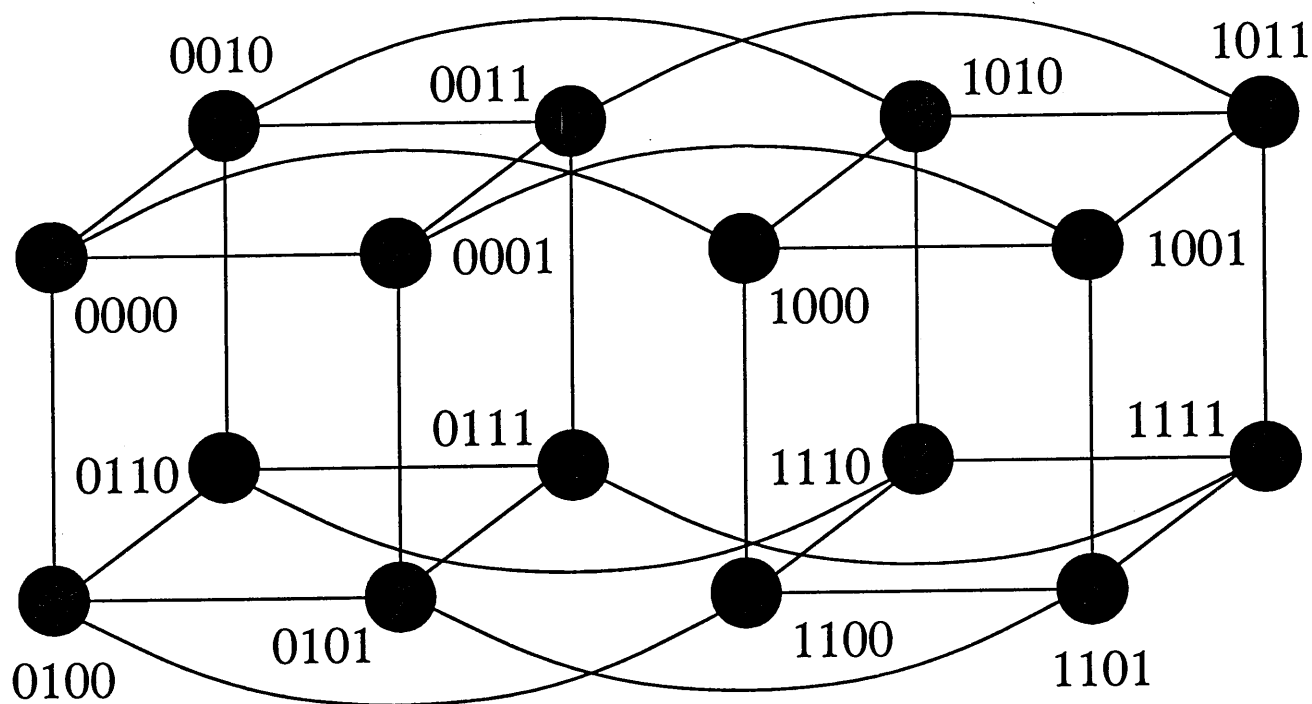


図 2.3: HC(2,4)

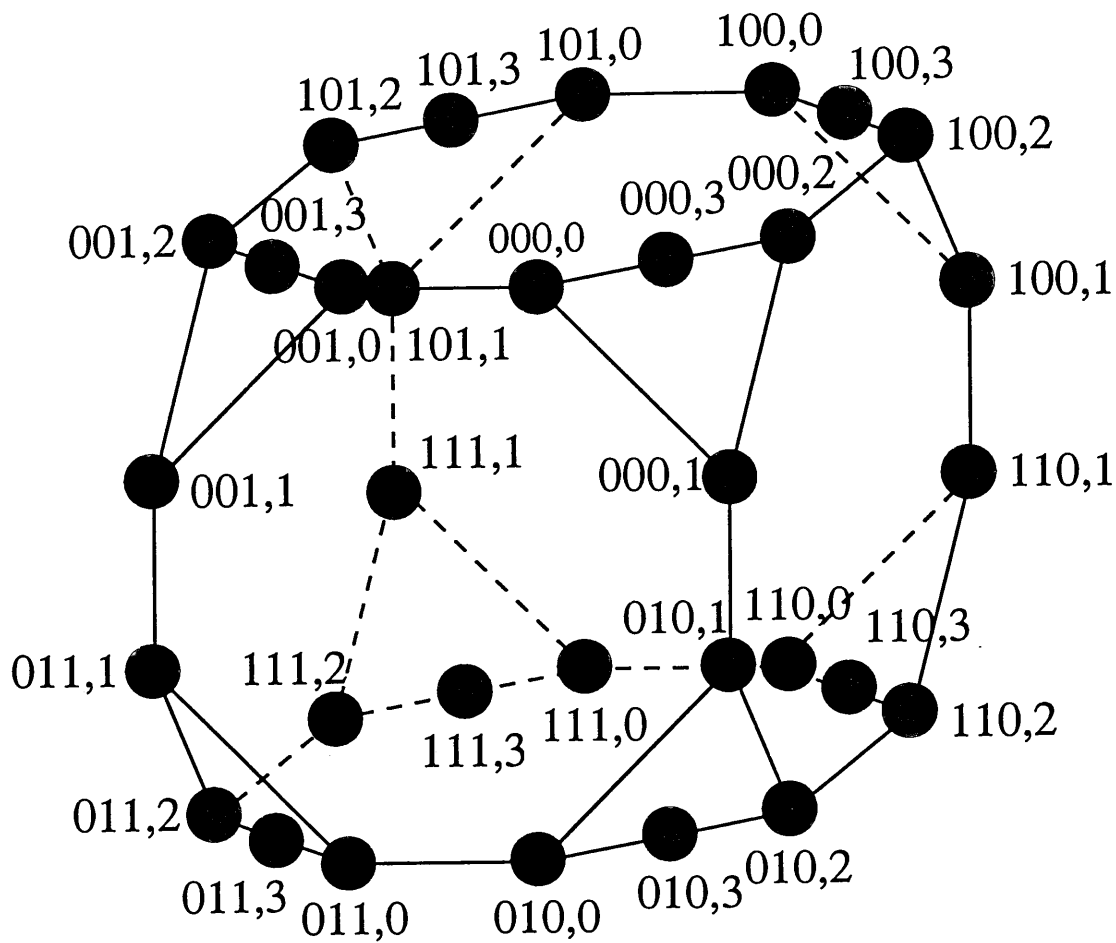


図 2.4: CCC(4,3)

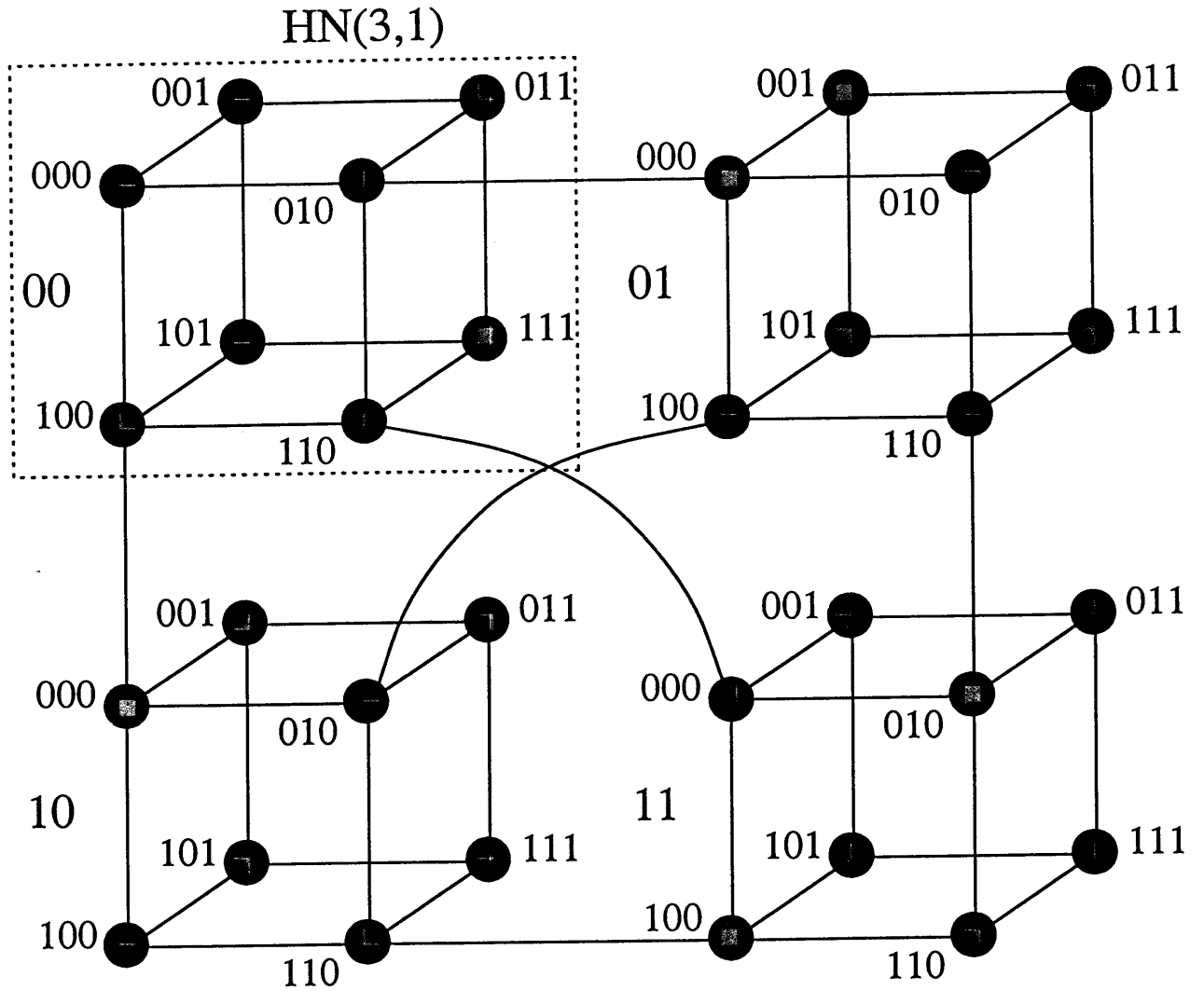


図 2.5: HN(3,2)

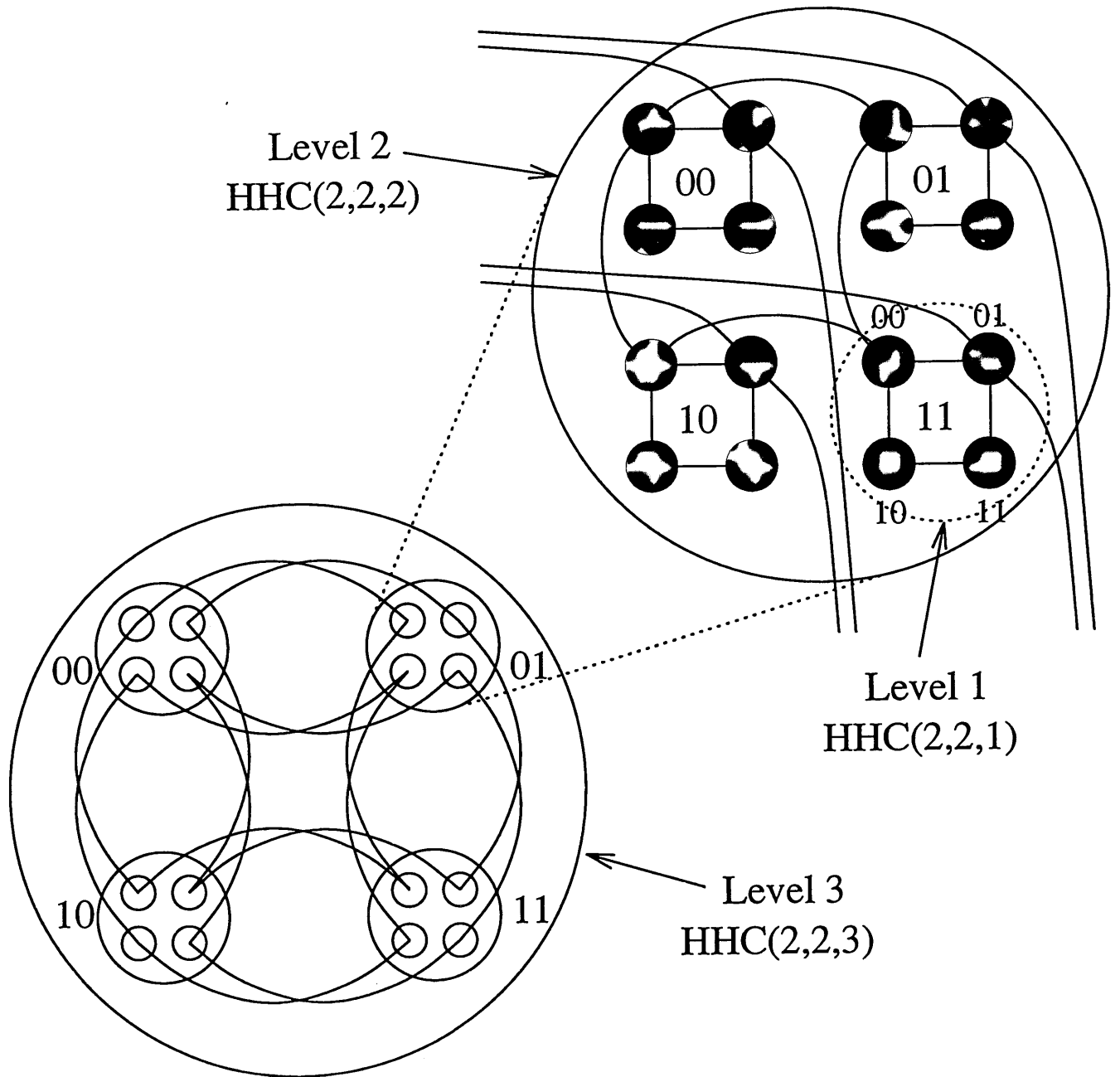


図 2.6: HHC(2,2,3)

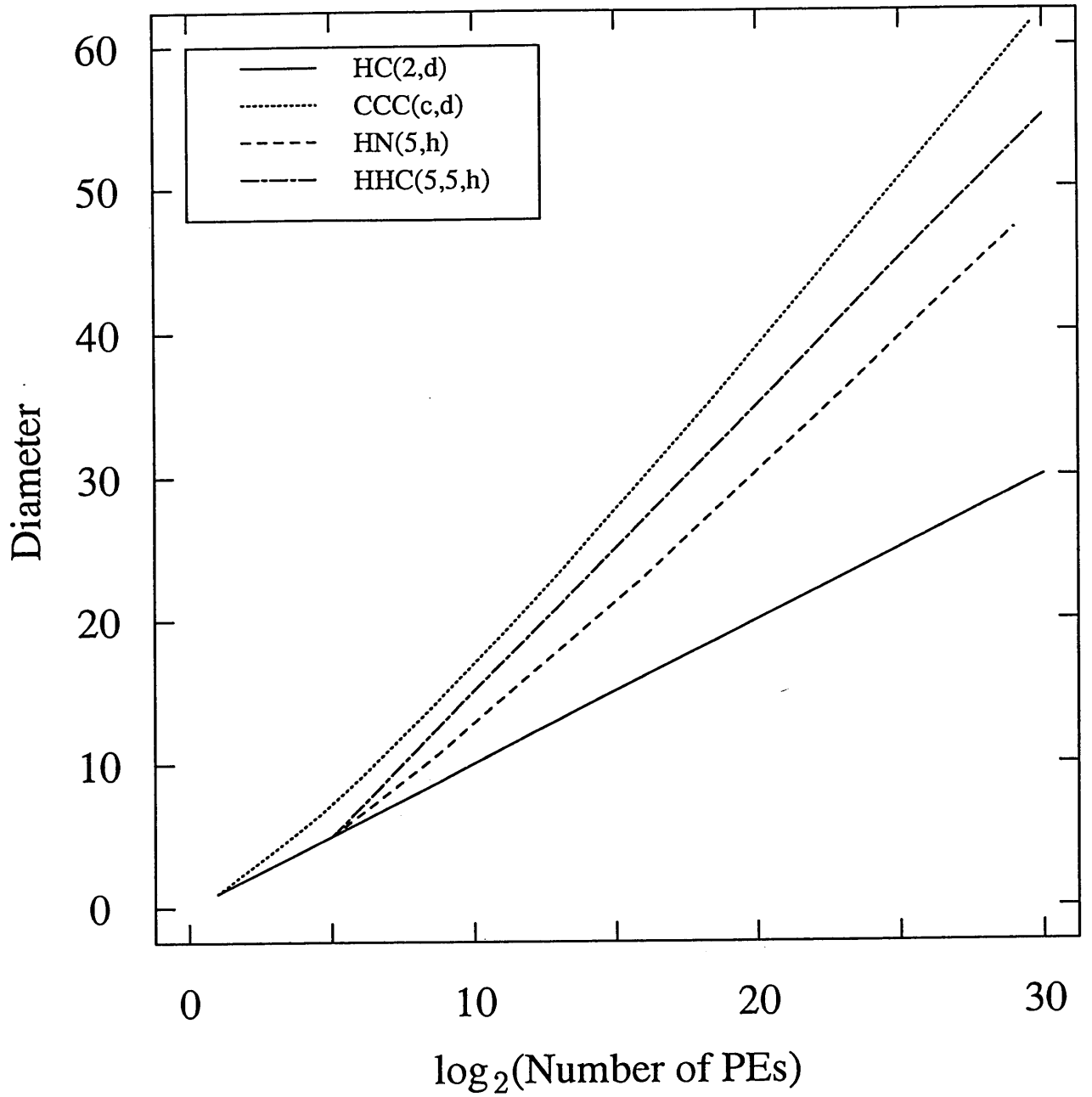


図 2.7: 直径の比較

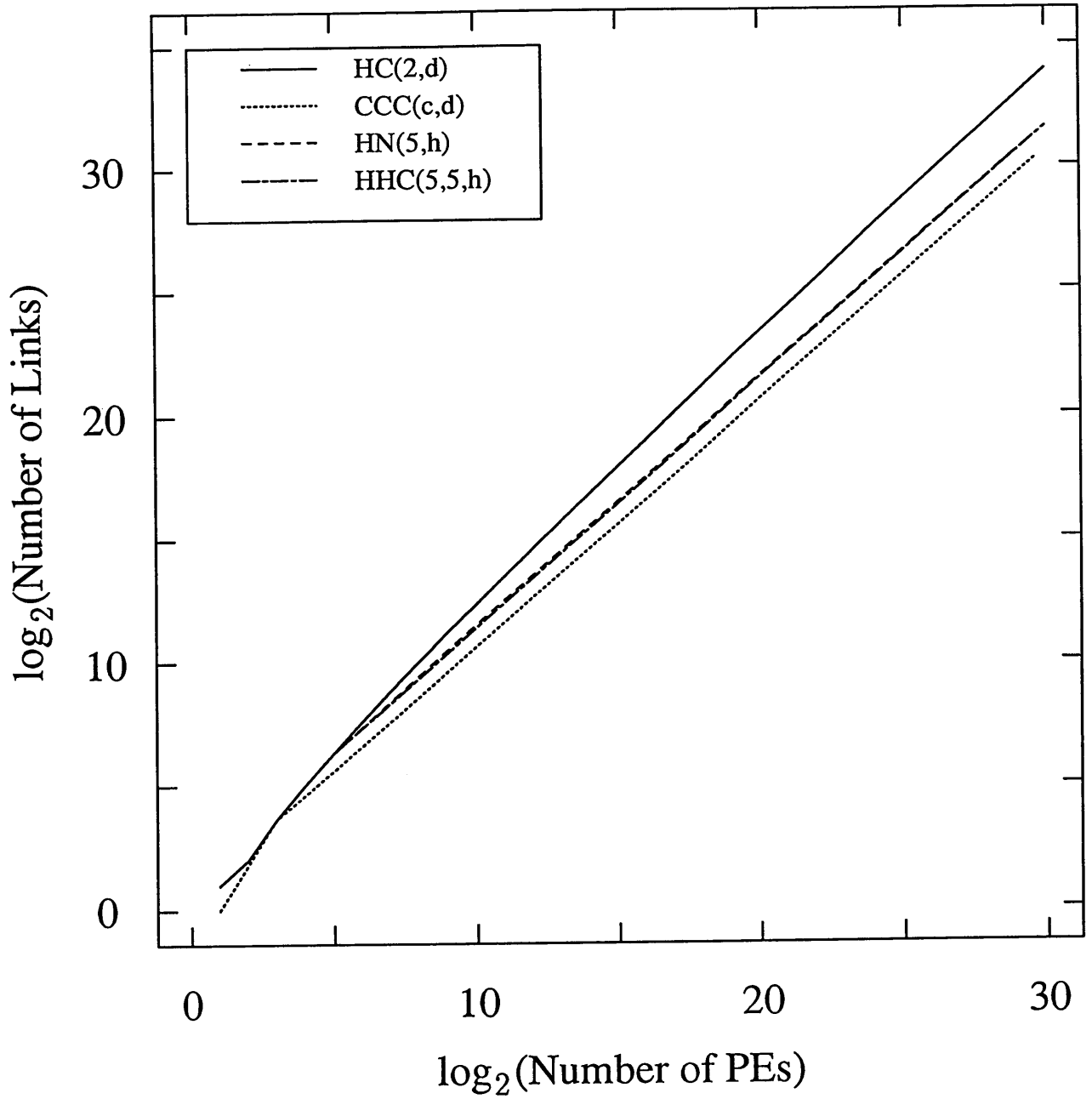


図 2.8: 総リンク数の比較

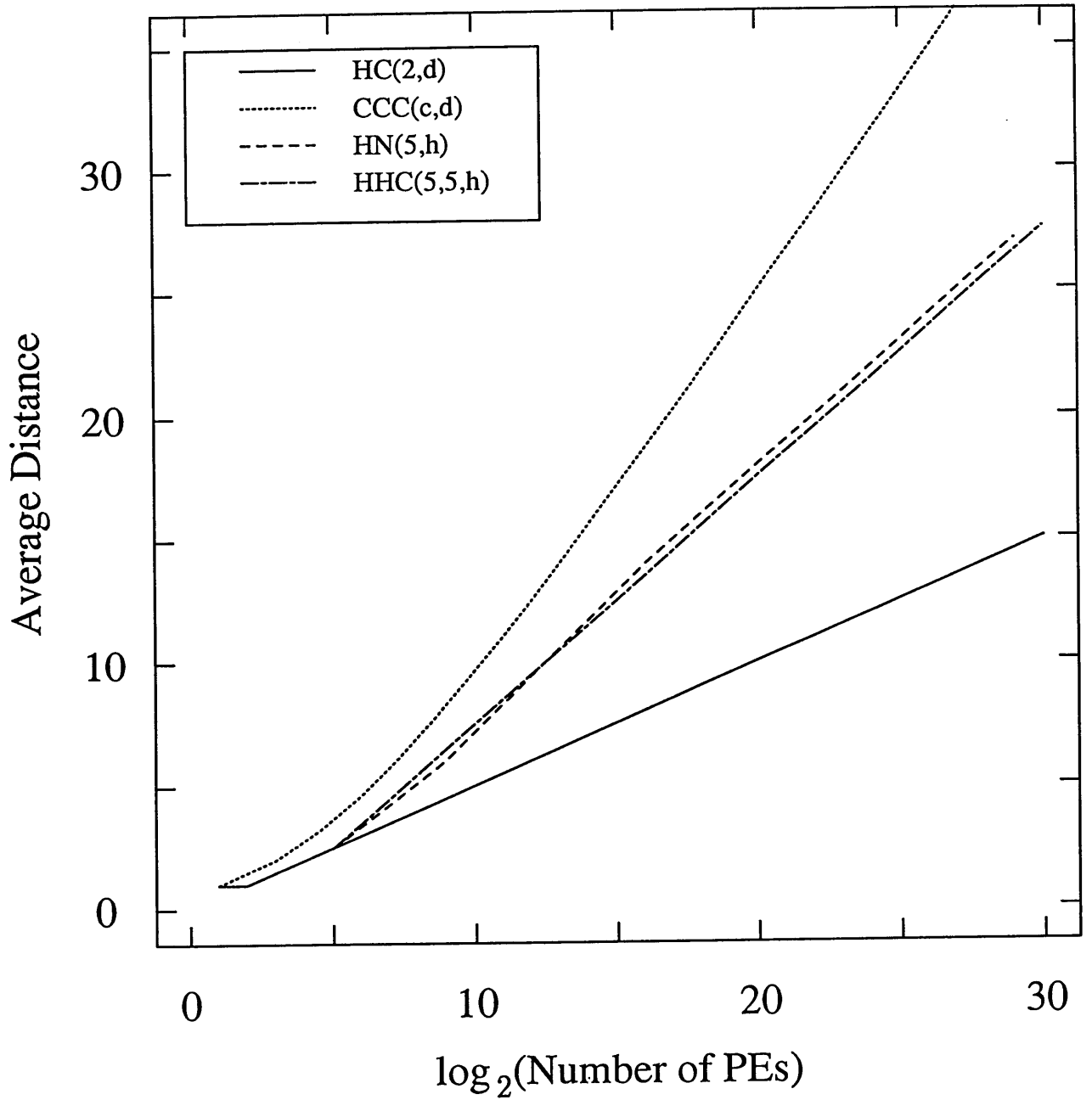


図 2.9: 平均距離の比較

第 3 章

ハイパキューブ結合型マルチプロセッサシステムのWSI構成法

3.1 はじめに

マルチプロセッサシステムのWSI構成法を考える場合、複雑なシステムのウェハ上のレイアウトと、製造過程においてウェハ上に発生する欠陥の救済法を考える必要がある。現在のVLSI回路では、配線領域として多階層を使用するものは一般的ではなく、本研究では、縦方向と横方向の2階層の配線領域を使用する構成法を考える。

2次元メッシュはネットワークのトポロジそのものが2次元であるので、平面上に容易にレイアウト可能であり、VLSI回路で実現しやすい。この2次元メッシュに関しては、これまでに様々なWSI構成法および再構成アルゴリズムが提案されている[4][5]。一方、本研究の対象としているハイパキューブ結合型のマルチプロセッサシステムはトポロジが2次元的ではないので、平面上のレイアウトが複雑となり、VLSI回路上では実現しにくい。しかし、第2章でも述べたように、ハイパキューブ結合型のマルチプロセッサシステムは超並列システムのプロセッサ間接続方式として有効である。したがって、小型で高速なシステムを作るためにも、そのWSI構成法の研究は重要となってくる。最近ではこのような複雑なシステムのWSI構成法の研究も行なわれるようになってきており、HCのWSI構成法については、伊藤らが詳しく検討している[6][7][8]。また、CCCについてもいくつかのWSI構成法が提案されている[16][17][18]。

本研究では基本のHCの構成法としては伊藤らの構成法を採用し、それを階層化させる方法を提案する。さらにこれを他のハイパキューブ結合型のシステムに適用させる。また、本研究ではPEやリンク、およびリンクの切り換えを行うスイッチなどの厳密な大きさや機能、実際の回路、および欠陥の検出回路などは考えないことにし、PE、冗長PE、および、リンク配線の配置のみを与える。

3.2節で実際の構成法と本構成法の各ネットワークへの適用法について述べる。

3.2 本研究で提案するWSI構成法

現在のVLSI回路の製造技術では回路の製造時の欠陥の発生が避けられない。したがって、1枚のウェハ上に巨大なシステムを搭載させるWSIシステムでは、通常回路の一部に冗長性を持たせて欠陥部分を回避してシステムを再構成する。

従来考えられてきた2次元メッシュをはじめとするマルチプロセッサシステムのWSI構成法の多くは、冗長なPEを付加してPE部分の欠陥にのみ対処するものであった。すなわち、リンクおよびスイッチなどの配線部分に発生した欠陥に対しては、そのリンク、あるいはスイッチの結合されているPEの欠陥と見なしたり、配線部分の欠陥は全く考えないことにするなど、詳しい考察を行っていないものが多かった。これは、1本のリンクや1個のスイッチの面積は1個のPEの面積に比べてはるかに小さく、これらの部分に欠陥が発生することはほとんど無いと考えられるためである。この考え方はPEの面積が大きくPE数が比較的少ない場合は成り立つ。しかし、本研究のように小規模のPEによる超並列システムのWSI構成法を考える場合は、この配線領域の面積が膨大となってきたこの部分の欠陥を無視できなくなってくる。従って、この配線部分の欠陥も救済できる構成法を考える必要がある。

冗長なリンク、およびスイッチなどを設けることにより、配線領域の欠陥に対処することはある程度可能である。ところが、冗長なリンクやスイッチを増やしてもクリティカルなリンクやスイッチは残り、その部分に欠陥が発生するとシステム全体が機能しなくなる。また、超並列システムのWSIシステムではリンク数およびスイッチ数が膨大となり、この方法では制御が非常に複雑となり、非現実的である。そこで、本研究では冗長構成を階層化させ配線領域の欠陥を上の方の階層の冗長回路で救済させる方法を探る。

具体的には2階層の冗長構成を考え、第1階層においてはPEを行方向と列方向の2次元メッシュ状に並べ、各行に冗長PEを付加して各行内で簡単な K -out-of- N 冗長構成を行う。さらに、第2階層では第1階層のブロックを行方向と列方向に2次元メッシュ状に並べて、各ブロック行に冗長なブロックを付加して、各ブロック行内でも K -out-of- N 冗長構成を行う。 K -out-of- N 冗長構成とは K 個の回路ユニットからなるシステムを構成させるために、 $(N-K)$ 個の冗長ユニットを付加し、合計 N 個のユニットから K 個の正常なユニットを用いてシステムを構成させる方法である。図3.1に2階層の冗長構成法のモデルを示す。

このように構成法を階層化させることにより、第1階層の非冗長部分(つまり、配線領域)の欠陥は第2階層の冗長ブロックを使用することにより救済することができる。また、欠陥の検出、および救済のプロセスを第1階層と第2階層に分けて欠陥の救済を局所化させることができる。なお、2次元メッシュ構造の階層型のWSI構成法はこれまでにいくつか研究されている[19][20]。

システムの再構成はウェハ上の回路の製造の最終段階に回路上の欠陥を検出した後に、リンク間の切り換えスイッチにより行う。また、このスイッチが動的に切り換え可能であるならば、システムの使用中のある段階に再構成を行うことも可能である。まず、第1階層のブロック内で冗長PEを使用して故障個所を避けてシステムを構成させる。次に、正常に動作しない第1階層のブロックを検出し、第2階層で冗長ブロックを使用して構成を行う。第1階層の各行内で冗長PEの数より多いPEが故障している場合や、本研究では詳しい考察を行っていないが、PEの故障として置き換えることができない配線部分に欠陥が発生している場合は、このブロックは正常に動作しないことになる。

3.2.1 ハイパキューブの構成法

HC(2,d)を構成させる場合、まず、第1階層のブロック内で d_1 次元のHCを構成させる。第2階層では第1階層で構成させた d_1 次元のHCのブロックを 2^{d_2} 個用いることにより全体を構成させる($d = d_1 + d_2$)。第1階層のブロック内(第2階層のブロック間)で $d_1(d_2)$ 次元のHCを構成させる場合のPE配列は、PE行内(行方向のブロック内)で $p_1(p_2)$ 次元のHCを構成させ、そのようなPE行(行方向のブロック)を $2^{d_1-p_1}$ 行($2^{d_2-p_2}$ ブロック行)用いる。

第1階層のブロック内の各PE行にはいくつかの冗長PEを加え、PE行内でK-out-of-N冗長構成を行う。さらに、第2階層のブロック行方向には冗長ブロックを加え、第2階層でもK-out-of-N冗長構成を行う。第1階層内の冗長PEを使用しても回避できない部分に欠陥が生じた場合、第2階層の冗長ブロックで置き換える。

図3.2にHC(2,d)のPE配置を示し、図3.3にHC(2,4)の構成例を示す。

図3.3で行方向と列方向のPE間の結合リンクはHC結合のリンク束であり、実際には n 次元のHC結合を行う場合は $2^n - 1$ 本の単一リンクから構成される。図3.4に3次元のHCのリンク束を示す。

また、図3.3のリンクの交点にある○で表した部分、すなわち、図3.4の□で囲んだ部分は行方向と列方向のHC結合のリンクをまとめて切り換えるスイッチである。このスイッチは図3.5に示すような様々な切り換えが可能であると仮定している。

n 次元のHC結合のリンクは、全ての部分で 2^{n-1} 本のリンク全部を使用するわけではない。しかし、超並列を目指すシステムの場合、リンク数が膨大となり、それを制御するスイッチの数も膨大となって実際の制御は困難となる。よってHC結合のリンク束を1本のHCリンクとしてまとめて制御を行うことにする。

本構成法でHCを構成するさいに必要なリンクの数は以下に示すようになる。

- 第1階層(d_1 次元HC。PE行内で p_1 次元のHCを構成させ、各PE行に sp_1 個の冗長PEを付加する。)

1. 行方向の p_1 次元HCリンク($2^{d_1-p_1}$ 本)

2. 列方向の $d_1 - p_1$ 次元HCリンク ($2^{p_1} + sp_1$ 本)
3. 2. をシフトするための行方向の $d_1 - p_1$ 次元HCリンク ($sp_1 \cdot (2^{d_1 - p_1} - 1)$ 本)
- 第2階層 (d_2 次元HC。ブロック行内で p_2 次元のHCを構成させ、各ブロック行に sp_2 個の冗長ブロックを付加する。)
 4. 行方向の p_2 次元HCリンク
($2^{d_1 - p_1} \cdot 2^{d_2 - p_2}$ 本)
 5. 列方向の $d_2 - p_2$ 次元HCリンク
($(2^{p_1} + sp_1) \cdot (2^{p_2} + sp_2)$ 本)
 6. 5. をシフトするための行方向の $d_2 - p_2$ 次元HCリンク
($sp_1 \cdot (2^{d_1 - p_1} - 1) \cdot 2^{d_2 - p_2}$ 本)
 7. 5. をブロックごとに、元の位置に戻すための行方向の $d_2 - p_2$ 次元HCリンク
($sp_1 \cdot 2 \cdot (2^{d_2 - p_2} - 1)$ 本)
 8. 5. をブロック間でシフトするための行方向の $d_2 - p_2$ 次元HCリンク
($sp_2 \cdot 2^{p_1} \cdot (2^{d_2 - p_2} - 1)$ 本)

HCの構成は第1階層と第2階層に分けて行う。第1階層内では1および2のリンクを用いて d_1 次元のHCを構成させる。欠陥等によりあるPEが正常に機能しない場合は、そのPEを故障PEとして使用しないことにし、各行内の冗長PEを含めた正常なPEのみを使用してリンクの結合を行う。1のリンクは行方向のHC結合リンクであり、各PEに結合されているリンクとの交点にあるスイッチを切り換える。また、2のリンクは列方向のHC結合を行うリンクであり、列方向のアドレスの位置が物理的にずれる部分では3のリンクを使用する。

第2階層の結合リンクのうち、4のリンクは1のリンクと同様に制御を行う。また、5と6のリンクはそれぞれ、2と3のリンクと同様である。第2階層でのブロック間の構成を簡単にするために、5のリンクは7のリンクを用いてブロックごとに同じ位置に戻すことにしている。また、8のリンクは正常に動作するブロックの位置が物理的にずれた場合に、5のリンクをシフトさせるためのリンクである。

3.2.2 巡回型ハイパキューブの構成法

CCC(c, d)を構成させる場合、 c 個のサイクルを1つのブロックとして第1階層の2行のPE行で構成させる。 c が奇数の場合は上の行のPEの数を1つ多くする。サイクル間の結合は d 次元のHC結合となるので、第2階層では第1階層のブロックを行方向に 2^p 個、列方向に 2^{d-p} 個並べて全体を構成させる。

第1階層のブロック内の各PE行と、第2階層の各ブロック行に冗長ブロックを加えて、それぞれの階層で K -out-of- N 冗長構成を行うことはHCの場合と同様である。

図3.6に $CCC(c, d)$ のPE配置を示し、図3.7に $CCC(4, 2)$ の構成例を示す。

第1階層のブロック内のサイクル内の結合リンクは単一のリンクであり、サイクル間の結合リンクはHCリンクとなる。

CCCを構成する場合に必要なリンク数は以下に示すようになる。

- 第1階層 (c 個のPEのサイクル。各PE行に sp_1 個の冗長PEを付加する。)
 1. 行方向のバイパスリンク (2 本)
 2. 列方向のバイパスリンク ($c/2 + sp_1 - 1$ 本)
 3. 列方向のサイクル間リンクをシフトするための行方向のリンク (sp_1 本)
- 第2階層 (d 次元HC。ブロック行内で p 次元のHCを構成させ、各ブロック行に sp_2 個の冗長ブロックを付加する。)
 4. 行方向の p 次元HCリンク
(2^{d-p} 本)
 5. 列方向の $d-p$ 次元HCリンク
($((d-p) + sp_1) \cdot (2^p + sp_2)$ 本)
 6. 5. をブロックごとに、元の位置に戻すための行方向の $d-p$ 次元HCリンク
($sp_1 \cdot 2 \cdot (2^{d-p} - 1)$ 本)
 7. 5. をブロック間でシフトするための行方向の $d-p$ 次元HCリンク
($sp_2 \cdot (d-p) \cdot (2^{d-p} - 1)$ 本)

第1階層内ではPE間にある行方向と列方向のリンクを用いてサイクル内の結合を行う。あるPEが故障している場合は、行方向のリンクに対するバイパス用リンク (1 および 2 のリンク) を用いてその故障PEをバイパスさせる。また、故障PEによって2行のPE行からなるサイクルの端のPEの位置がずれる場合には、3のリンクを用いて列方向のリンクをシフトさせる。

第2階層は d 次元のHC結合となる。4と5はそれぞれ行方向と列方向のHC結合のリンクである。HCの場合と異なり、第1階層のブロックは2行のPE行からなるので、5のリンクをシフトさせるための行方向のHCリンクは必要ない。また、6は5のリンクをブロックごとに元の位置に戻すためのHCリンクであり、7はブロック間の結合リンクをシフトさせるためのHCリンクである。5および7のリンクは、 d および p のとり方によっては使用しないリンクがある。

3.2.3 ハイパネット

HN(d, h)を構成させる場合、クラスタ(基本ネット)である d 次元のHCを1つのブロックとして第一階層で構成させる。第2階層ではこのようなブロックを行方向に q 個、列方向に C/q 個並べて全体を構成させる。ここで C はクラスタの数であり、 $C = 2^{2^{h-1} \cdot (d-2) + h+1-d}$ である。

HC、およびCCCの場合と同様に、第1階層および第2階層でそれぞれ K -out-of- N 冗長構成を行う。

図3.8にHN(d, h)のPE配置を示し、図3.9にHN(2,3)の構成例を示す。

第1階層のクラスタ内の結合リンクはHCリンクあり、第2階層のクラスタ間の結合リンクは単一リンクの結合である。

HNを構成する場合に必要なリンク数は以下に示すようになる。

- 第1階層(d 次元HC。PE行内で p 次元のHCを構成させ、各PE行に sp_1 個の冗長PEを付加する。)
 1. 行方向の p 次元HCリンク (2^{d-p} 本)
 2. 列方向の $d-p$ 次元HCリンク ($2^p + sp_1$ 本)
 3. 2.をシフトするための行方向の $d-p$ 次元HCリンク ($sp_1 \cdot (2^{d-p} - 1)$ 本)
- 第2階層(C 個のブロック。ブロック行内で q 個のブロックを配置させ、各ブロック行に sp_2 個の冗長ブロックを付加する。)
 4. ブロック間結合用の行方向の単一リンク ($(h-1) \cdot (C/q)$ 本)
 5. ブロック間結合用の列方向の単一リンク ($(h-1) \cdot (q-1)$ 本)
 6. ブロック内PE結合用の列方向の単一リンク ($2^{d-p} \cdot (2^p + sp_1) \cdot (q + sp_2)$ 本)

第1階層内は d 次元のHCであり、HCの第1階層内の結合と同様である。

第2階層は C 個のブロック内での完全結合となる。4、5、および6のリンクは単一リンクである。4のリンクは行方向のブロック間の結合用のリンクであり、5のリンクは5のリンクを各ブロック行間で結ぶ列方向のリンクである。また、6のリンクは各PEと4のリンクを結ぶリンクである。

3.2.4 階層型ハイパキューブ

HHC(d_1, d_2, h)を構成させる場合、クラスタである d_1 次元のHCを第1階層で構成させる。クラスタから見るとHHC(d_1, d_2, h)は $(h-1) \cdot d_2$ 次元のHCとなるので、第2階層で

は第1階層のブロックを行方向に $2^{p_2 \cdot d_2}$ 個、列方向には $2^{(h-1-p_2) \cdot d_2}$ 個並べて全体を構成させる。

HC、CCC、およびHNの場合と同様に、第1階層および第2階層でそれぞれ K -out-of- N 冗長構成を行う。

図3.10に $\text{HHC}(d_1, d_2, h)$ のPE配置を示し、図3.11に $\text{HHC}(2, 2, 2)$ の構成例を示す。
HHCを構成する場合に必要なリンク数は以下に示すようになる。

- 第1階層 (d_1 次元HC。PE行内で p_1 次元のHCを構成させ、各PE行に sp_1 個の冗長PEを付加する。)
 1. 行方向の p_1 次元HCリンク ($2^{d_1-p_1}$ 本)
 2. 列方向の $d_1 - p_1$ 次元HCリンク ($2^{p_1} + sp_1$ 本)
 3. 2. をシフトするための行方向の $d_1 - p_1$ 次元HCリンク ($sp_1 \cdot (2^{d_1-p_1} - 1)$ 本)
- 第2階層 ($(h-1) \cdot d_2$ 次元HC。ブロック行内で $p_2 \cdot d_2$ 次元のHCを構成させ、各ブロック行に sp_2 個の冗長ブロックを付加する。)
 4. 行方向の $p_2 \cdot d_2$ 次元HCリンク
($((h-2)/2^{p_1} + 1) \cdot 2^{(h-1-p_2) \cdot d_2}$ 本)
 5. 列方向の $(h-1-p_2) \cdot d_2$ 次元HCリンク
($(2^{p_1} + sp_1) \cdot (2^{p_2 \cdot d_2} + sp_2)$ 本)
($h-1 < 2^{p_1}$ のときは、 $((h-1) + sp_1) \cdot (2^{p_2 \cdot d_2} + sp_2)$ 本となる。)
 6. 5. をシフトするための行方向の $(h-1-p_2) \cdot d_2$ 次元HCリンク
($sp_1 \cdot ((h-2)/2^{p_1}) \cdot 2^{(h-1-p_2) \cdot d_2}$ 本)
 7. 5. をブロックごとに、元の位置に戻すための行方向の $(h-1-p_2) \cdot d_2$ 次元HCリンク
($sp_1 \cdot 2 \cdot (2^{(h-1-p_2) \cdot d_2} - 1)$ 本)
 8. 5. をブロック間でシフトするための行方向の $(h-1-p_2) \cdot d_2$ 次元HCリンク
($sp_2 \cdot 2^{p_1} \cdot (2^{(h-1-p_2) \cdot d_2} - 1)$ 本)
($h-1 < 2^{p_1}$ のときは、 $sp_2 \cdot (h-1) \cdot (2^{(h-1-p_2) \cdot d_2} - 1)$ 本となる。)

第1階層内は d_1 次元のHCであり、HCの第1階層内の結合と同様である。また、第2階層は $(h-1) \cdot d_2$ 次元のHCであり、HCの場合と同様であるが、階層数 h によって4、5、6、および8のリンクは必要なリンク数が少なくなる。4および6において、 $(h-2)/2^{p_1}$ は整数の商であり、 $(h-2)/2^{p_1+1}$ はブロック間のHC結合を行うブロック内のPE行の数を表す。つまり、4のリンクは行方向のブロック間のHC結合を行うリンクであり、ブロック間のHC結合を行うPE行においてのみ必要である。6のリンクはブロック間のHC結

合を行うPE行の間のみで必要である。また、5および8のリンクは $(h-1) < 2^{p_1}$ の場合、つまりブロック間の結合が行われているPEの数がブロック内の1行のPEの数よりも少ない場合には、ブロック間の結合が行われるPEとして使用される可能性のあるPEの数だけ用意すればよい。

3.3 まとめ

本章では本研究で提案するハイパキューブ結合型マルチプロセッサシステムのWSI構成法について述べた。その構成法はネットワークを2階層に分けてそれぞれの階層で K -out-of- N 冗長構成を行うものであった。2階層の冗長構成にすることにより、第1階層の配線部分に発生した欠陥を第2階層の冗長構成により救済することが可能となる。また、ハイパキューブ結合の結合リンクを1本のリンクとして扱うことにより、制御するスイッチの数を大幅に減らすことができる。

次章ではこの構成法の評価を行う。

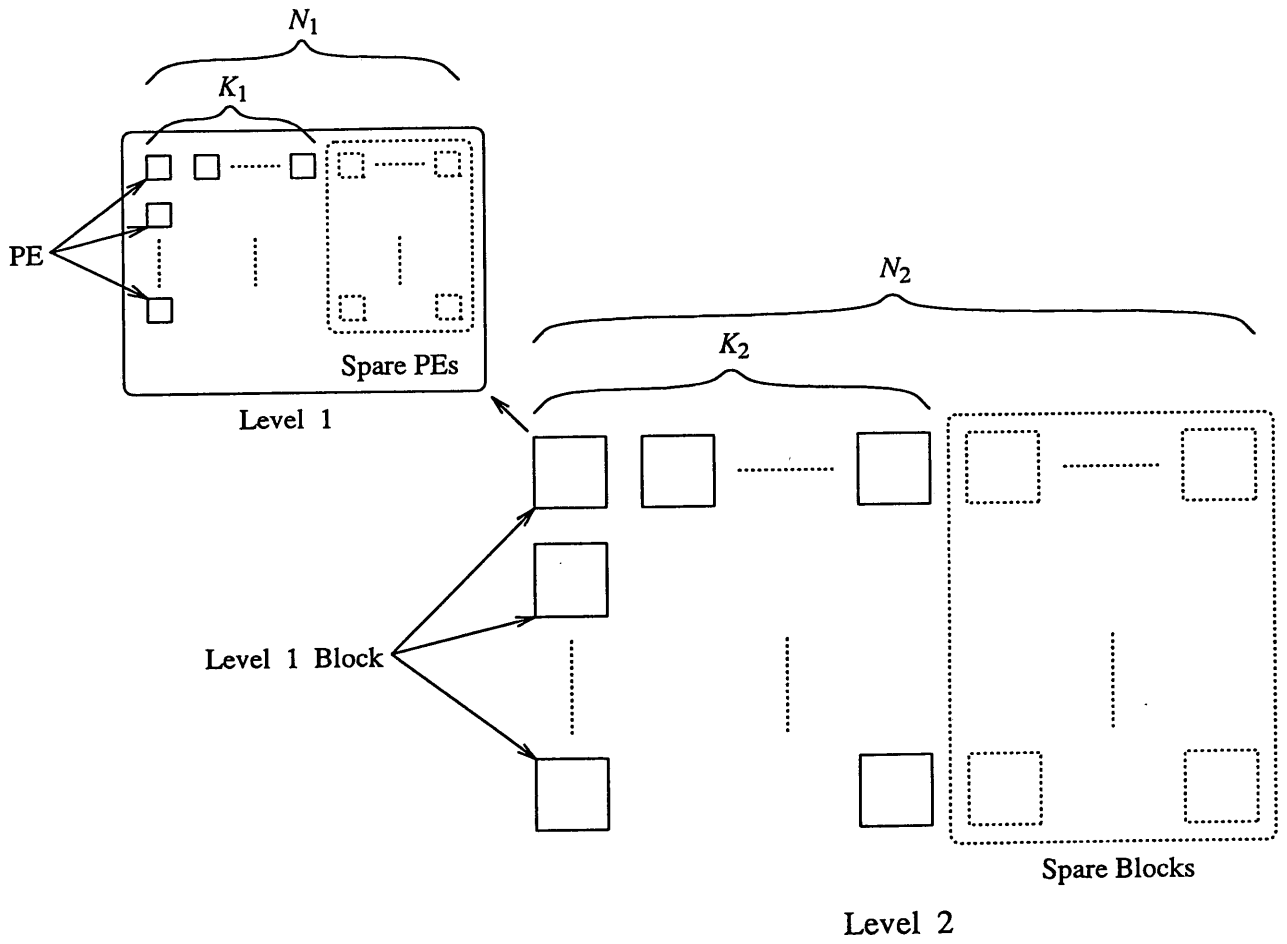


図 3.1: 階層型冗長構成法

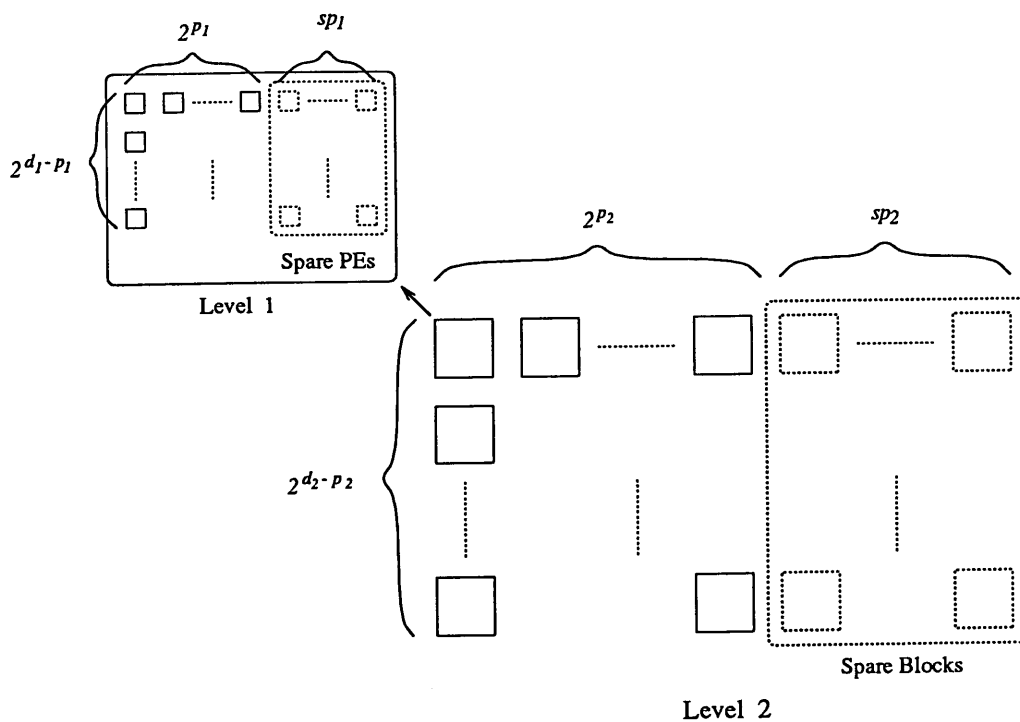


図 3.2: $HC(2, d)$ の PE 配置

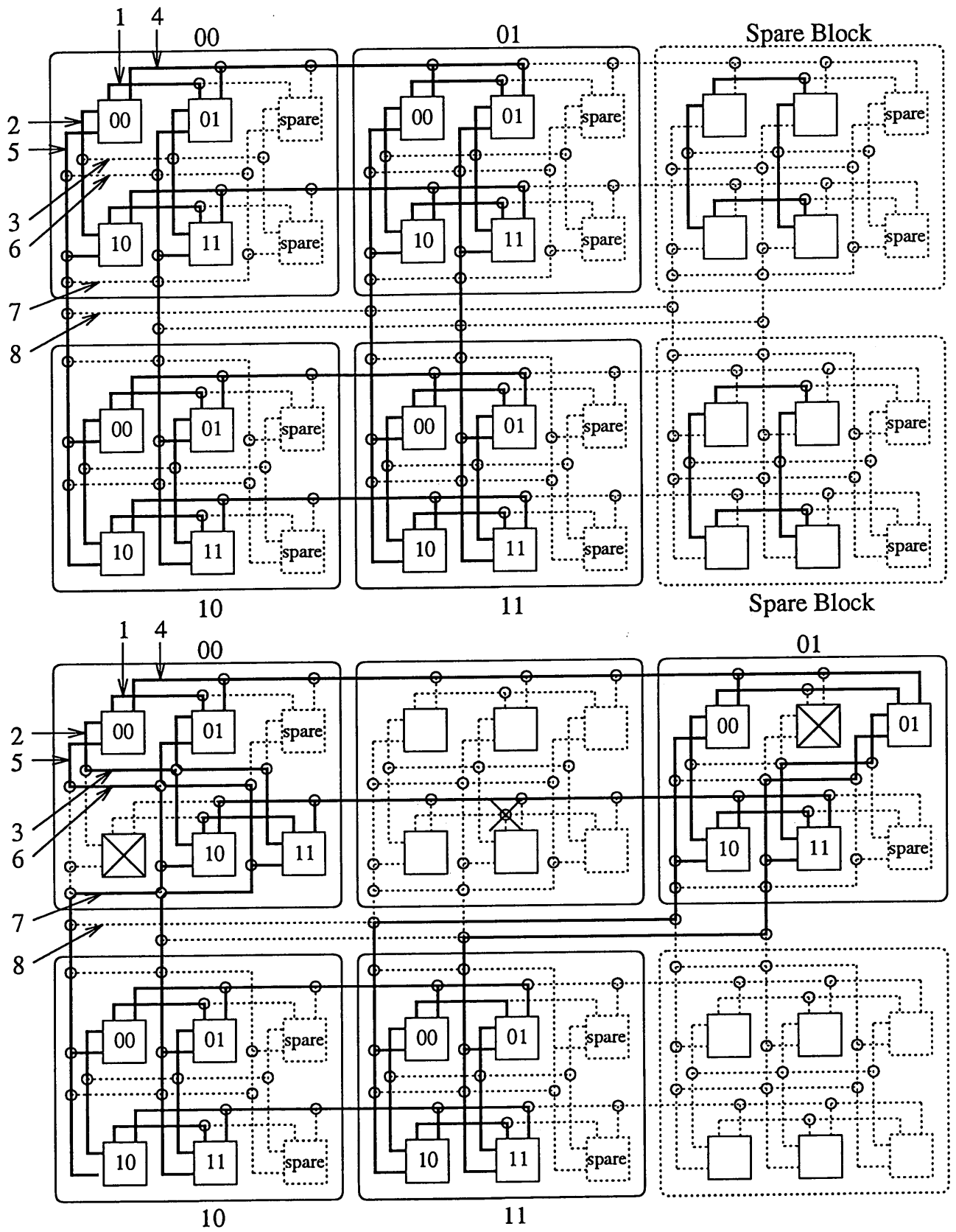


図 3.3: HC(2,4) のWS I 構成例

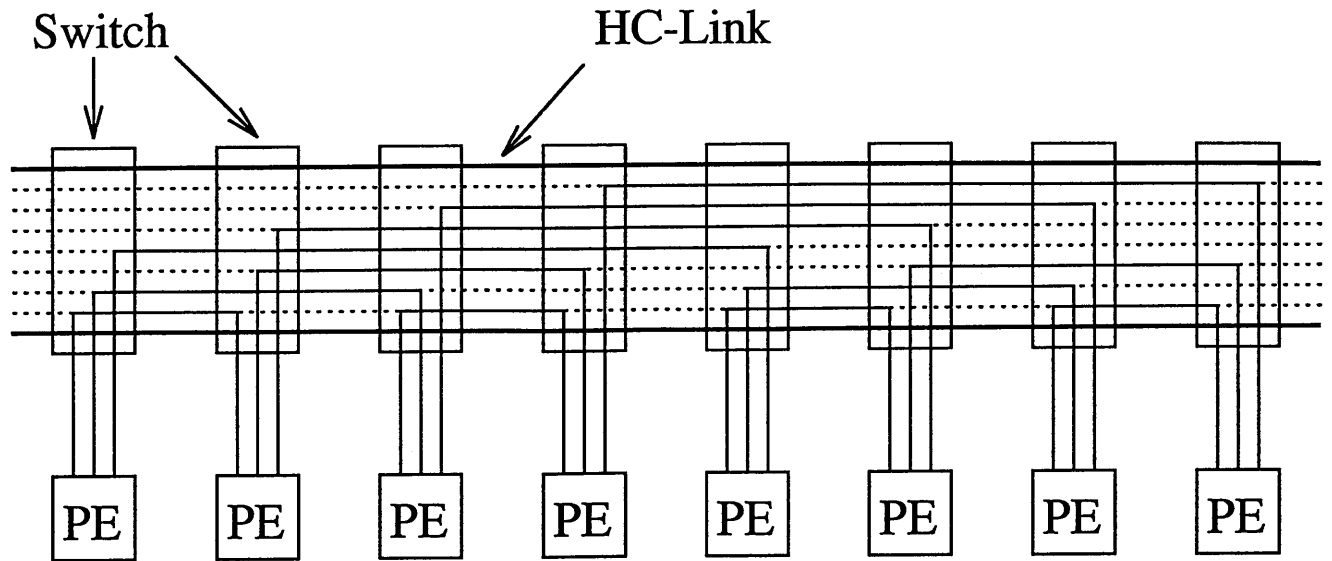


図 3.4: HC結合リンク束

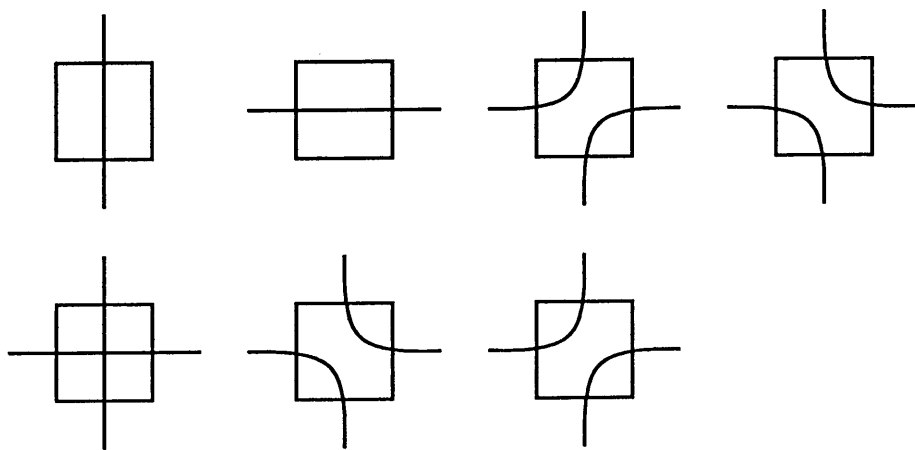


図 3.5: スイッチの切り換えパターン

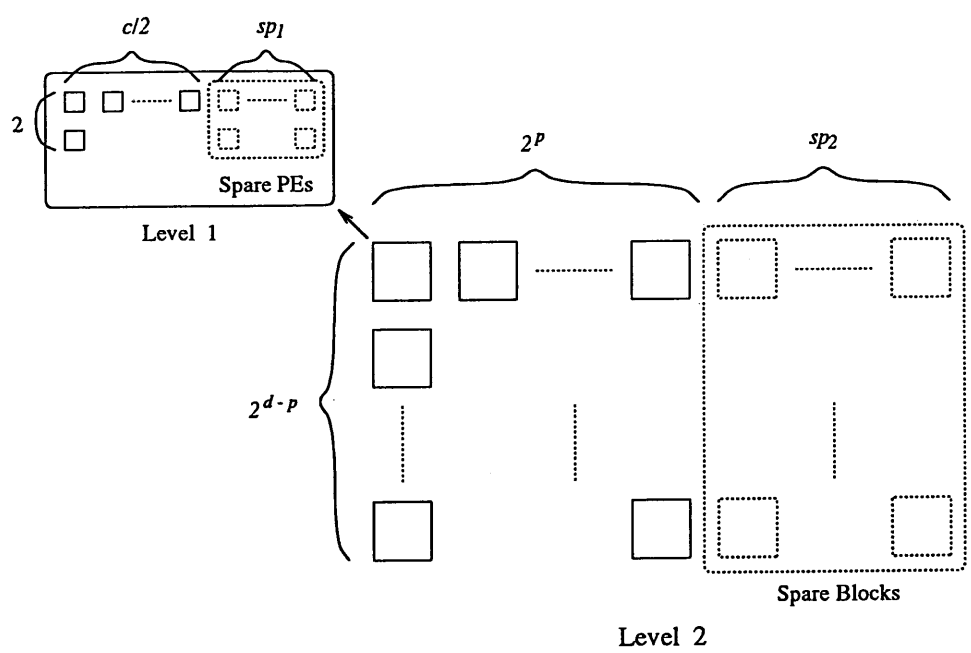


図 3.6: CCC(c, d) の P E 配置

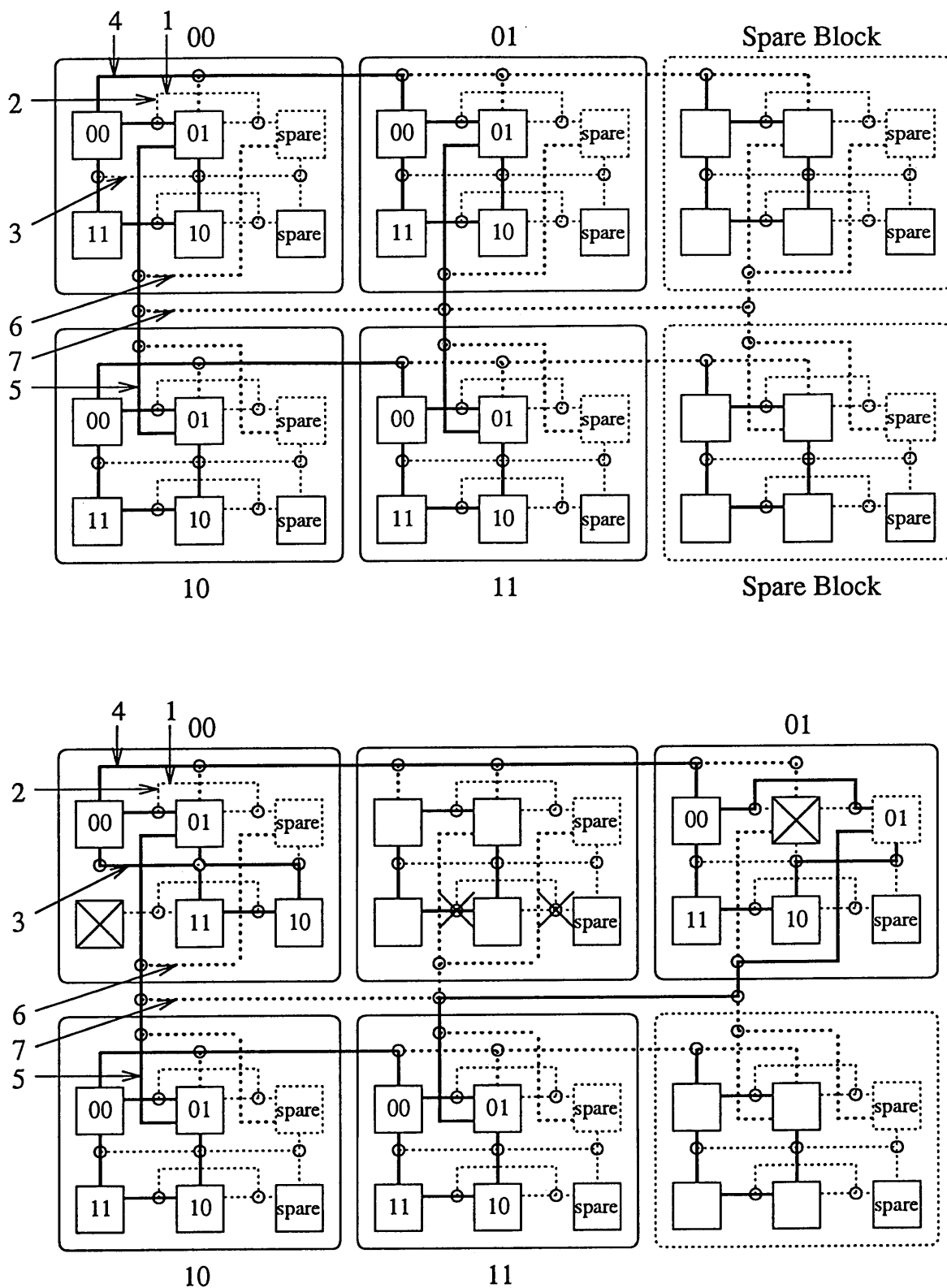


図 3.7: CCC(4,2) のWSI 構成例

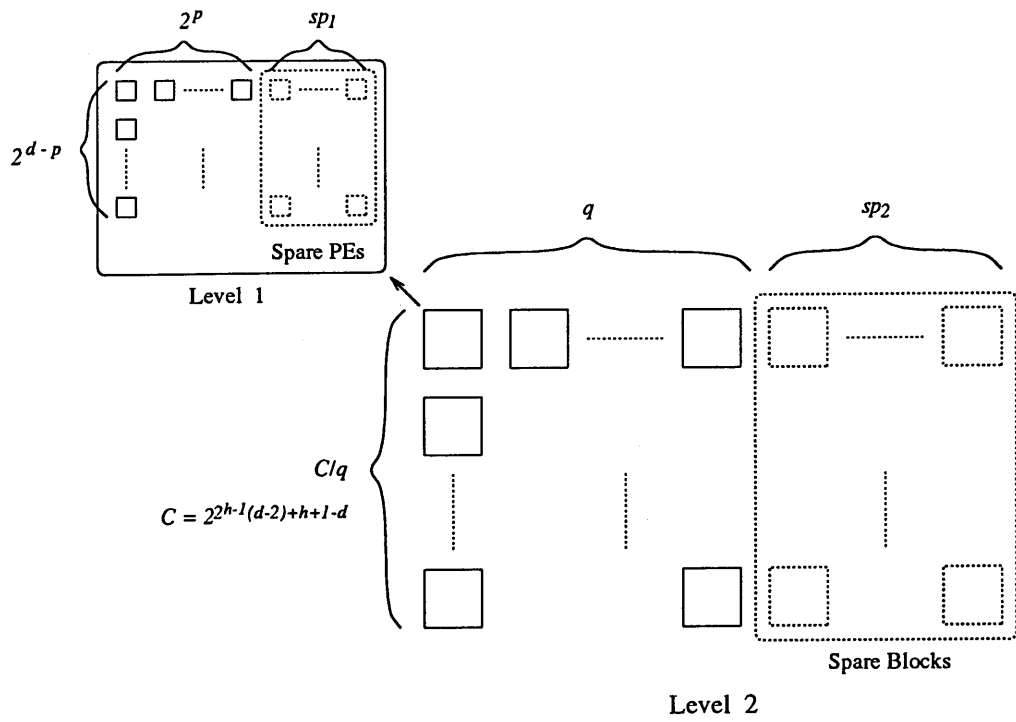


図 3.8: $HN(d, h)$ の PE 配置

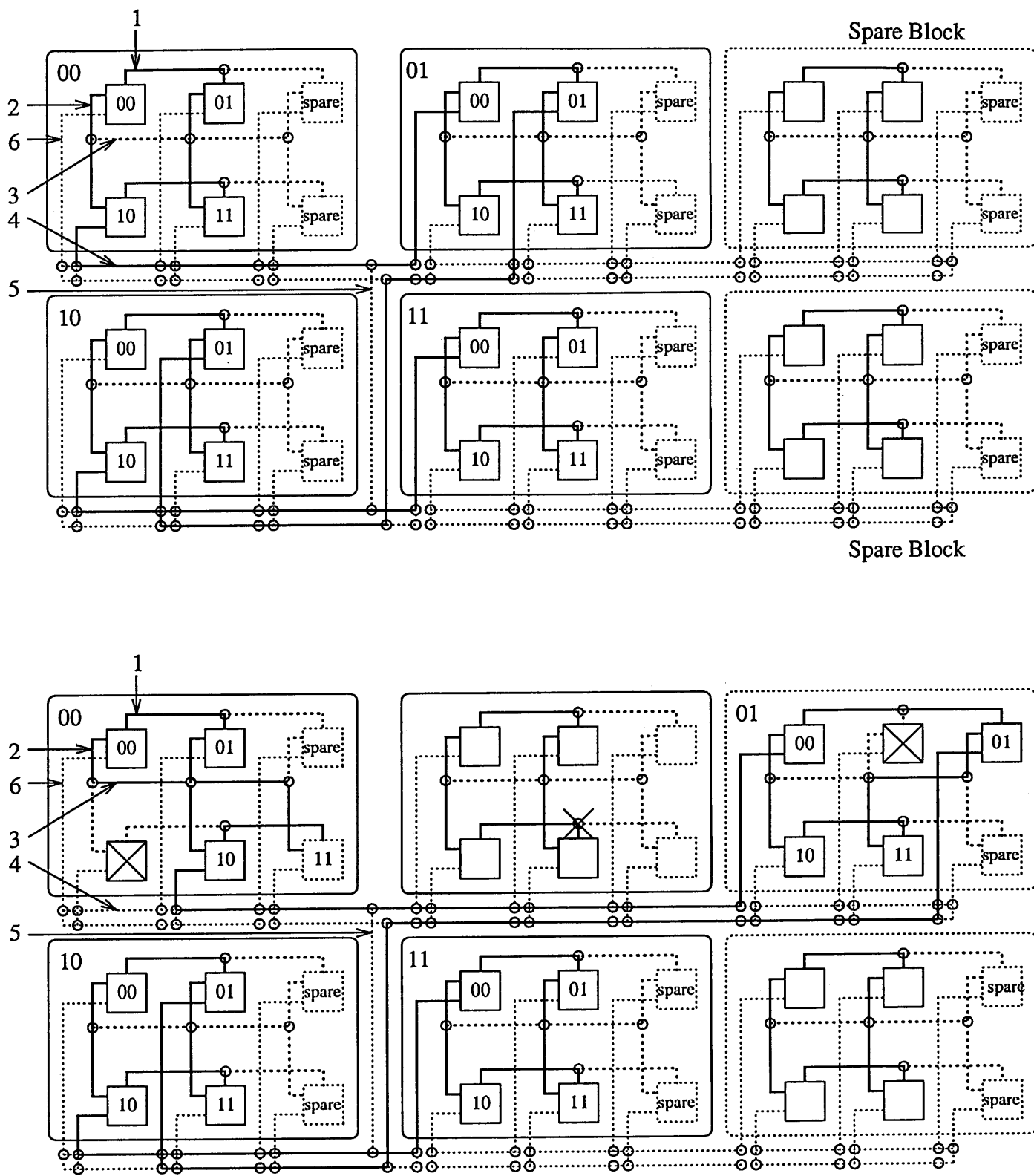


図 3.9: HN(2,3) のWSI 構成例

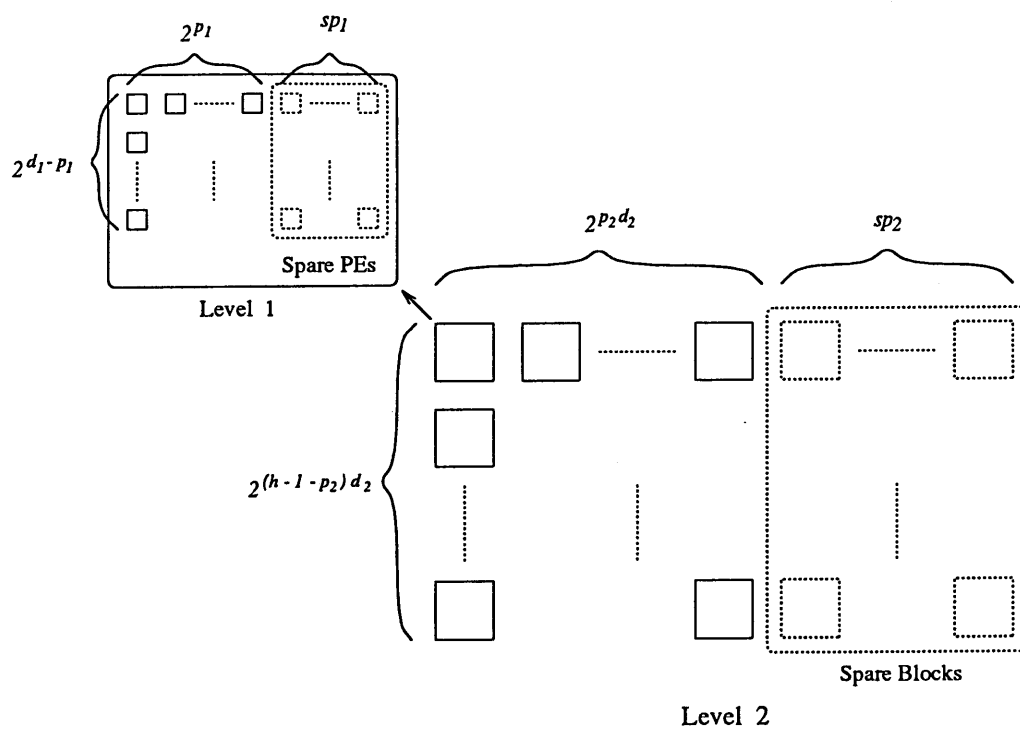


図 3.10: HHC(d_1, d_2, h) の PE 配置

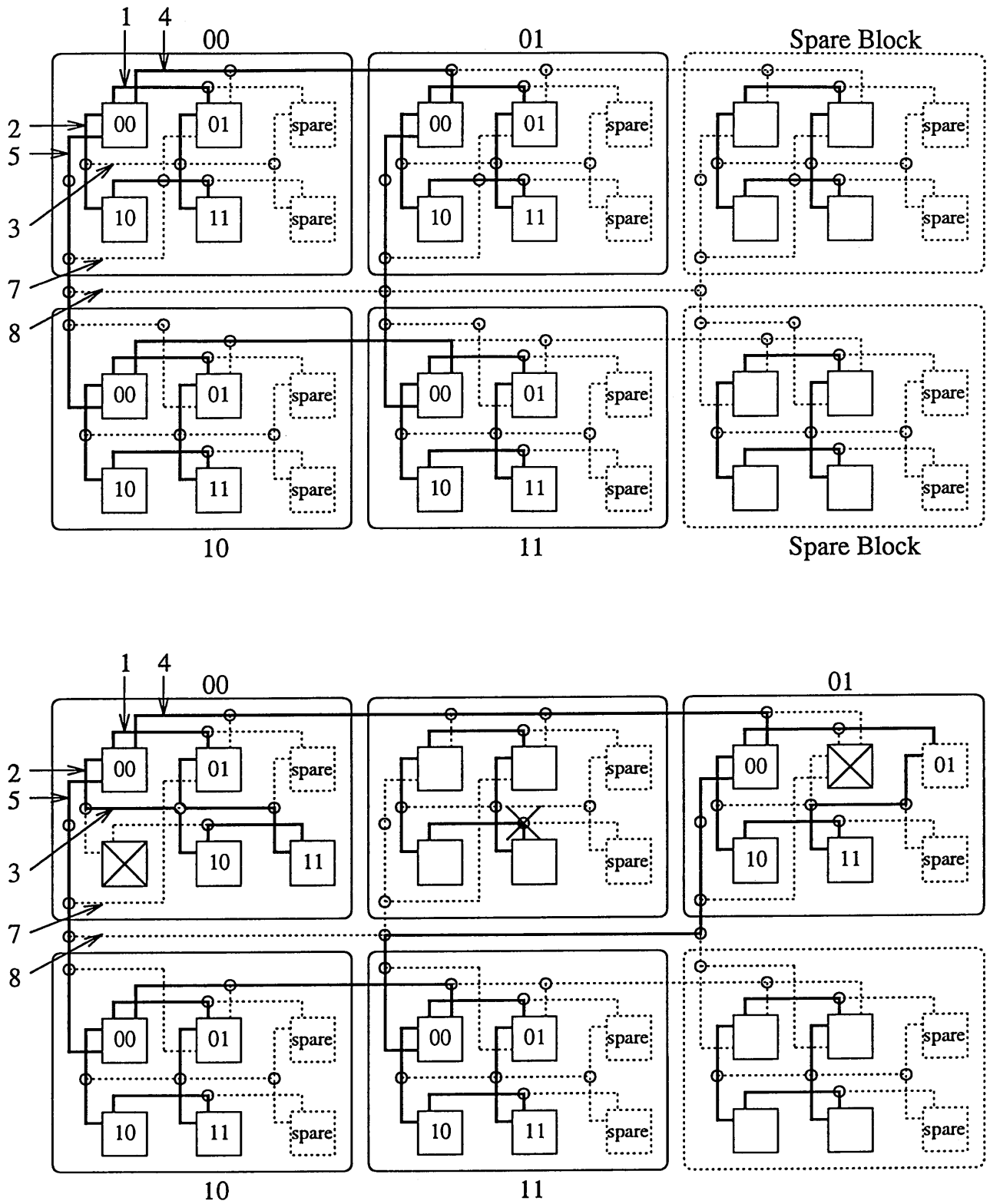


図 3.11: HHC(2,2,2) のWSI 構成例

第 4 章

WSI システムの性能評価

4.1 はじめに

本章では第3章で提案したWSI構成法の評価を行う。

WSIシステムの性能評価としては、レイアウト面積、歩留まり、最大リンク長、および切り換えスイッチ数などが挙げられる。本章ではこれらのうちのレイアウト面積と歩留まりについて考察を行う。4.2節でレイアウト面積について述べ、4.3節で歩留まりについて述べる。

4.2 レイアウト面積

レイアウト面積とは、PE、冗長PE、冗長ブロック、リンク、スイッチ、および空白領域の面積を含めた全面積を表す。本研究では回路の最小線幅を $1.0\mu\text{m}$ とし、1本のリンク幅を W_{link} 、 n 次元のHC結合のリンク幅を $(2^n - 1) \cdot W_{link}$ 、PEを幅が W_{pe} の正方形としてレイアウト面積を計算する。なお、切り換えスイッチの大きさはリンクの幅の中に埋め込めるものと仮定している。

第3章で述べたハイパキューブ結合型のマルチプロセッサシステムのレイアウト面積を、3.2節で示した構成法の行方向と列方向のリンク数を考慮して求める。 A_1 を第1階層のブロック内の全PEの面積とブロック内の結合リンク部分の面積を合わせた面積とし、 A_2 を全ブロックの面積とブロック間の結合リンク部分の面積を合わせた面積(全レイアウト面積)とすると、 A_1 および A_2 は以下の式のように表される。

$$\begin{aligned} A_1 &= (PE_{r1} \cdot W_{pe} + Link_{c1} \cdot W_{link}) \cdot (PE_{c1} \cdot W_{pe} + Link_{r1} \cdot W_{link}) \\ A_2 &= A_1 \cdot Block_r \cdot Block_c \\ &\quad + (Link_{r2} \cdot W_{link}) \cdot \{PE_{r1} \cdot Block_r \cdot W_{pe} + (Link_{c1} \cdot Block_r + Link_{c2}) \cdot W_{link}\} \end{aligned} \quad (4.1)$$

$$\begin{aligned}
 &+(Link_{c2} \cdot W_{link}) \cdot \{PE_{c1} \cdot Block_c \cdot W_{pe} + (Link_{r1} \cdot Block_c + Link_{r2}) \cdot W_{link}\} \\
 &-(Link_{r2} \cdot W_{link}) \cdot (Link_{c2} \cdot W_{link})
 \end{aligned} \tag{4.2}$$

$$\left(\begin{array}{l}
 PE_{r1} : \text{第1階層のブロック内の冗長PEを含めた1行のPE数} \\
 PE_{c1} : \text{第1階層のブロック内の1列のPE数} \\
 Link_{r1} : \text{第1階層のブロック内の結合を行うための行方向のリンク数} \\
 Link_{c1} : \text{第1階層のブロック内の結合を行うための列方向のリンク数} \\
 Block_r : \text{第2階層の冗長ブロックを含めたブロック行のブロック数} \\
 Block_c : \text{第2階層のブロック列のブロック数} \\
 Link_{r2} : \text{第2階層のブロック間の結合を行うための行方向リンク数} \\
 Link_{c2} : \text{第2階層のブロック間の結合を行うための列方向のリンク数}
 \end{array} \right)$$

ここで、 $PE_{r1}, PE_{c1}, Link_{r1}, Link_{c1}, Block_r, Block_c, Link_{r2}, Link_{c2}$ を各ネットワークについてまとめると表4.1のようになる。

1個のPEの面積を A_{pe} とすると A_{pe} にはプロセッサ、演算ユニット、通信制御回路、およびメモリなどの面積が含まれている。マルチプロセッサシステムのPEにおいては、通信制御回路部分の面積は、そのPEに接続されているリンクの数によってかなり異なってくる[21]。そこで、本研究ではPEの面積を通信制御回路部分の面積と他の部分の面積に分け、通信制御回路の面積をそのPEに接続されているリンク数に応じて変化させることにする。

具体的には、1本のリンクの制御に必要な回路の面積を A_{com} 、通信制御回路以外の面積を A_{other} として、PEの面積を次式で求める。

$$A_{pe} = W_{pe}^2 = D_i \cdot A_{com} + A_{other} \tag{4.3}$$

ここで D_i は各PEのリンク接続数である。本研究では各ネットワークのPEは全て同じ通信制御回路を持つとして、 D_i は各ネットワークの次数とする。

以下では、 $W_{link} = 4 \cdot 2.0 \mu\text{m}$ および $A_{other} = 1\text{mm}^2$ とする。HC、CCC、HN、およびHHCの各ネットワークについて、同じ数のPEからなるシステムを構成するときのレイアウト面積を計算する。このとき、各ネットワークの構成法に対して全冗長PE数が同じになるようにPEの配置を行なうものとし、そのときのPE配列を表4.2に示す。また、HC、HN、HHCは全く同じPE配列をとっている。

$A_{com} = 0.1\text{mm}^2$ として通信制御回路の面積を考慮したときの各ネットワークのレイアウト面積の比較を、図4.1に示す。グラフの横軸は冗長PEと冗長ブロックを含めないときのPE数(つまり構成させるシステムの全PE数)を示している。また、縦軸は冗長PE、および冗長ブロックを含めた各WSI構成法の全レイアウト面積を表している。 2^{16} 個のPEからなるシステムを構成させるために必要なレイアウト面積は、HCで約 10000cm^2 以上、CCC、HN、HHCでは約 1000cm^2 以上となっている。面積の違いを明らかにするために、HCのレイアウト面積を1としたときの他のネットワークのレイアウト面積の割合を

図4.2に示す。レイアウト面積は小さい順に、CCC、HN、HHC、HCの順となる。また、構成するネットワークのサイズを大きくするにつれて、HCに対するレイアウト面積の割合は小さくなっている。これは第2章の図2.8のグラフに示されているように、これらのシステムがHCに比べて、総リンク数を減らした構造であることと、構成するシステムが大きくなるときの総リンク数の増加の割合がHHCの方大きいからである。このことから、超並列システムの構成を目指す場合、CCC、HN、およびHHCのハイパキューブ結合型マルチプロセッサシステムはHCに比べて実現しやすいといえる。

次に、 $A_{com} = 0\text{mm}^2$ のとき、つまり各PEの面積を各ネットワークの次数によらず同じとしたときのレイアウト面積をの比較を図4.3に示す。この場合は各ネットワークのPEの面積は同じで、冗長なPE数も同じであるので、純粹にリンクおよびスイッチ部分の面積の違いとなる。この場合は小さい順に、HN、CCC、HHC、HCの順になっている。また、CCCとHHCのレイアウト面積はほとんど同じである。HNが最小となったのは、HNは第2階層のブロック間の結合が単一のリンクからなる結合であり、ブロック間のハイパキューブ結合を行うCCC、HHCよりもこの部分の面積が小さくなっているからと考えられる。

$A_{com} = 0.1\text{mm}^2$ の場合と $A_{com} = 0\text{mm}^2$ の場合の実際の面積を表4.3に示す。また、これらのレイアウト面積の差をグラフにしたものを図4.4に示す。この表、およびグラフは、それぞれのサイズのシステムについて、CCCが最もレイアウト面積の差が小さくなっていることを示している。これは、CCCの次数がネットワークのサイズによらず3で他のネットワークよりも小さくなっているからである。

$A_{com} = 0\text{mm}^2$ (通信制御回路の面積を考慮しない場合)はHNが最もレイアウト面積が小さくなり、 $A_{com} = 0.1\text{mm}^2$ として、通信制御回路の面積を考慮に入れると、次数の最も小さいCCCのレイアウト面積が最小となることから、配線領域が小さく、次数が少ないネットワークがWSI構成に適しているということになる。

4.3 歩留まり

レイアウト面積は各ネットワークの構成法の実際のレイアウトの面積を表しているだけで、WSIシステムの特徴である冗長構成による欠陥の救済の評価とはならない。そこで、1枚のウェハ上でシステムを構成したときにそれが正常に動作するかどうかを表す確率である歩留まり(Yield)を考える。本来歩留まりは統計値であるが、ここでは欠陥の分布を仮定して確率的に求める。歩留まりが1とは、ウェハ上で構成したシステムが必ず動作することを示している。また、歩留まりはシステム構成(確)率と呼ぶこともできる。

現在の集積回路の製造技術では、ウェハ上の回路の製造過程において様々な欠陥が生じる。この欠陥の分布がポアソン分布(Poisson distribution)に従ってに発生する点状のランダム欠陥と考えると、歩留まり Y は、回路面積 A とウェハ上の平均の欠陥数を表す欠陥密

度 D を用いて、

$$Y = e^{-AD} \quad (4.4)$$

で表される。したがって、ウェハ上のシステムの歩留まりはシステムを構成するために必要な面積の増加とともに急激に低下し、欠陥救済機能のない大規模なシステムでは歩留まりはほとんど0となる。よって、W S I システムでは冗長構成による欠陥救済が不可欠となっている。また、歩留まりは冗長構成により変わってくる。

ウェハ上の欠陥は実際にはクラスタ (集団) 状に発生することが分かっている。そこで、ウェハ上の欠陥がクラスタ状に分布していると仮定して歩留まりを計算する。本研究では欠陥の分布が負の2項分布 (negative binomial distribution) に従うモデルを用いて歩留まりを計算する [22][23]。この方法はV L S I、あるいはW S I 回路の歩留まりの計算に一般的に用いられている方法であり、その計算方法を本節の最後に示す。

欠陥の分布が負の2項分布に従うとした場合、歩留まり Y は回路面積 Y と欠陥密度 D を用いて、

$$Y = (1 + AD/\alpha)^{-\alpha} \quad (4.5)$$

で表される。ここで、 α は回路の製造過程におけるクラスタ係数であり、 α が大きいということはクラスタが小さいことを示し、 α が小さいということはクラスタが大きいことを示している。

各ネットワークの歩留まりの計算は階層構造を考慮して行う。各階層でレイアウト面積を冗長部分とそれ以外の部分 (スイッチおよびリンク部分) に分ける。冗長部分の歩留まりに関しては、1つのP Eの歩留まり (Y_{pe}) (第1階層)、あるいは1つのブロックの歩留まり (Y_1) (第2階層) からそれぞれの階層の冗長構成を考慮して計算する。一方、スイッチおよびリンク部分の歩留まりに関しては、この部分の合計の面積中に欠陥が1つも存在しない確率として計算する。

$$Y = Y(\text{リンクおよびスイッチ部分}) * Y(\text{冗長部分}) \quad (4.6)$$

第1階層の歩留まりを Y_1 、第2階層の歩留まり (つまり全体の歩留まり) を Y_2 とすると Y_1, Y_2 は以下のように求められる。

$$Y_1 = Y(A_{link_1}) \cdot \left[\sum_{k=0}^{K_1} \{ P E_{r_1} C_k \cdot Y_{pe}^{P E_{r_1} - k} \cdot (1 - Y_{pe})^k \} \right]^{P E_{c_1}} \quad (4.7)$$

$$Y_2 = Y(A_{link_2}) \cdot \left[\sum_{k=0}^{K_2} \{ B l o c k_r C_k \cdot Y_1^{B l o c k_r - k} \cdot (1 - Y_1)^k \} \right]^{B l o c k_c} \quad (4.8)$$

$$\left(\begin{array}{l} Y(A_{link_1}): \text{第1階層の配線部分の歩留まり} \\ Y(A_{link_2}): \text{第2階層の配線部分の歩留まり} \\ K_1: \text{第1階層での冗長P Eを含まないP E数} \\ K_2: \text{第2階層での冗長ブロックを含まないブロック数} \end{array} \right)$$

ここで、 $Y_{pe}, Y(A_{link_1}), Y(A_{link_2}), A_{link_1}, A_{link_2}$ は以下の式から求められる。

$$Y_{pe} = (1 + W_{pe}^2 D / \alpha)^{-\alpha} \quad (4.9)$$

$$Y(A_{link_1}) = (1 + A_{link_1} D / \alpha)^{-\alpha} \quad (4.10)$$

$$Y(A_{link_2}) = (1 + A_{link_2} D / \alpha)^{-\alpha} \quad (4.11)$$

$$A_{link_1} = A_1 - Link_{r_1} \cdot Link_{c_1} \cdot W_{link}^2 \quad (4.12)$$

$$A_{link_2} = A_2 - A_1 \cdot Block_r \cdot Block_c \quad (4.13)$$

以下では $W_{link} = 4 \cdot 2.0 \mu m$, $A_{com} = 0.1 mm^2$, $A_{other} = 1 mm^2$, $\alpha = 2.0$ とする。本研究で提案した W S I 構成法のリンク配線は、端から端までの全てを使用しているわけではなく、全く使用していない部分に欠陥が生じる場合もある。従って、ここでは P E 部分の欠陥密度を $D = 2.0 / cm^2$ 、リンクおよびスイッチ部分の欠陥密度を $D = 1.0 / cm^2$ として歩留まりを求めることにする。

通信制御回路を考慮した場合の各構成法の歩留まりの比較を図 4.5 に示す。構成させるシステムのサイズ、冗長 P E の配置の仕方、および各 P E のリンク接続数などの条件はレイアウト面積の計算の場合と同じとしている。

この条件で歩留まりを計算すると構成させるシステムのサイズが 2^8 では、歩留まりは、H C で約 0.28、H H C で約 0.67、C C C で約 0.76、H N で約 0.85 となった。したがって、C C C、H N、H H C は H C に対してかなり歩留まりを上げることができるといえる。

全く同じ条件でレイアウト面積を計算した場合 (図 4.1) では、C C C と H N のレイアウト面積は C C C の方がわずかに小さかった。一方、歩留まりは H N の方がかなり大きくなった。これは、H N の第 2 階層間のブロック間結合のリンクの面積が C C C に比べて小さく、この部分に対する歩留まりが効いてくるためと考えられる。

また、P E 数が 2^{12} 個になると、全てのネットワークで歩留まりは 0 となる。これは P E 数が多くなると第 1 階層内、および第 2 階層内でのリンク配線部分の面積が大きくなり、この部分に対する歩留まりがほとんど 0 となるからである。従って、本構成法によりシステムを構成させる場合、 $W_{link} = 4 \cdot 2.0 \mu m$, $A_{com} = 0.1 mm^2$, $A_{other} = 1 mm^2$ の条件ならば、P E 数が 2^{10} 程度の中程度のマルチプロセッサシステムまでなら構成可能であるといえる。

欠陥の分布が負の 2 項分布に従う場合の歩留まりの計算法

- (I) 1 枚のウェハを n 個の部分に分け、各部分に欠陥が存在する確率が等しく p であるとすると、 n 個の内の x 個が欠陥を含んでいる確率は 2 項分布に従う。

$$P(X = x) = {}_n C_x \cdot p^x \cdot (1 - p)^{n-x} \quad (4.14)$$

- (II) p が非常に小さく、 $np = \lambda$ を一定として、 n を非常に大きくすると、この確率分布は

ポアソン分布に従う。

$$P(X = x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda} \quad (4.15)$$

(III) 実際には欠陥はクラスタ状に発生するので、 p はウェハの各部分によって異なる。そこで λ がガンマ分布 $f(\lambda)$ に従って分布すると仮定する。

$$f(\lambda) = \frac{1}{\gamma^\alpha \Gamma(\alpha)} \cdot \lambda^{\alpha-1} \cdot e^{-\lambda/\gamma} \quad (4.16)$$

(IV) (II) の λ を (III) の分布に従って平均化させると、欠陥の分布は負の2項分布となる [24]。(ポアソン分布の mixture)

$$\begin{aligned} P(X = x) &= \int_0^\infty \frac{\lambda^x}{x!} \cdot e^{-\lambda} \cdot f(\lambda) d\lambda \\ &= \frac{1}{x! \Gamma(\alpha) \gamma^\alpha} \int_0^\infty \lambda^{x+\alpha-1} \cdot e^{-(1+1/\gamma)\lambda} d\lambda \\ &= \frac{1}{x! \Gamma(\alpha) \gamma^\alpha} \cdot \frac{\Gamma(x+\alpha)}{(1+1/\gamma)^{x+\alpha}} \\ &= \frac{\Gamma(x+\alpha)}{x! \Gamma(\alpha)} \cdot \frac{\gamma^x}{(1+\gamma)^{x+\alpha}} \\ &= {}_{x+\alpha-1}C_x \cdot \left(\frac{\gamma}{1+\gamma}\right)^x \left(\frac{1}{1+\gamma}\right)^\alpha \end{aligned} \quad (4.17)$$

(V) α をウェハの製造過程におけるクラスタ係数、 A をウェハの面積、 D をウェハの欠陥密度とすると、 γ は次式で表される。

$$\gamma = \frac{AD}{\alpha} \quad (4.18)$$

歩留まりをウェハ上に欠陥が全く存在しない確率として (IV) より求める。

$$Y = P(X = 0) = \left(1 + \frac{AD}{\alpha}\right)^{-\alpha} \quad (4.19)$$

4.4 まとめ

本章では第3章で述べたハイパキューブ結合型のマルチプロセッサシステムの W S I 構成法に関する評価法について述べ、評価を行った。本研究では P E やリンクの大きさを仮定してレイアウト面積と歩留まりについて評価を行った。P E の面積を数 mm^2 程度、リンクの幅を $4.2.0\mu\text{m}$ とした場合でも、P E 数が 2^{10} 個を超えると歩留まりはほとんど0になってしまう。これは、超並列を目指したシステムの場合 P E 間の結合を行なうリンクの数が膨大となり、この部分に対する歩留まりがほとんど0になってしまうからである。実際に

は、この配線領域には欠陥が発生してもシステムの動作には関係ない領域がかなりあると考えられるが、本研究では配線領域の細かい構成、および細かい評価は行っていない。したがって、さらに厳密な評価を行うためには、配線部分の細かい構成法の検討、およびその評価方法の検討が必要になってくる。

なお、本研究では構成法を2階層とすることにより、第1階層の配線部分の欠陥を第2階層の冗長構成で救済する方法を採用したが、この階層化の評価は行わなかった。簡単に考察を行うと、2階層の階層型の構成法では第2階層の冗長単位の歩留まりが第1階層の冗長単位(P E)の歩留まりよりも高くなれば階層化の効果が得られる。しかし、第2階層において、第1階層のブロック間の結合に使用する配線面積が大きいと、この部分に欠陥が生じた場合は正常に動作しなくなる。したがって、HNのように、第2階層における第1階層のブロック間の配線部分の面積が小さくなる場合に、特に階層構成は有効であると考えられる。

表 4.1: 各ネットワークのPE数とリンク数

	HC(2, d)	CCC(c, d)
PE_{r1}	$2^{p_1} + s_1$	$c/2 + s_1$
PE_{c1}	$2^{d_1-p_1}$	2
$Link_{r1}$	$(2^{p_1} - 1) \cdot PE_{c1}$ $+ (2^{d_1-p_1} - 1) \cdot s_1 \cdot (PE_{c1} - 1)$	$2 + s_1$
$Link_{c1}$	$(2^{d_1-p_1} - 1) \cdot PE_{r1}$	$PE_{r1} - 1$
$Block_r$	$2^{p_2} + s_2$	$2^p + s_2$
$Block_c$	$2^{d_2-p_2}$	2^{d-p}
$Link_{r2}$	$(2^{p_2} - 1) \cdot PE_{c1} \cdot Block_c$ $+ (2^{d_2-p_2} - 1) \cdot s_1 \cdot (PE_{c1} - 1) \cdot Block_c$ $+ (2^{d_2-p_2} - 1) \cdot s_1 \cdot 2 \cdot (Block_c - 1)$ $+ (2^{d_2-p_2} - 1) \cdot s_2 \cdot 2^{p_1} \cdot (Block_c - 1)$	$(2^p - 1) \cdot Block_c$ $+ (2^{d-p} - 1) \cdot s_1 \cdot 2 \cdot (Block_c - 1)$ $+ (2^{d-p} - 1) \cdot s_2 \cdot (d - p) \cdot (Block_c - 1)$
$Link_{c2}$	$(2^{d_2-p_2} - 1) \cdot PE_{r1} \cdot Block_r$	$(2^{d-p} - 1) \cdot ((d - p) + s_1) \cdot Block_r$

	HN(d, h)	HHC(d ₁ , d ₂ , h)
PE_{r1}	$2^p + s_1$	$2^{p_1} + s_1$
PE_{c1}	2^{d-p}	$2^{d_1-p_1}$
$Link_{r1}$	$(2^p - 1) \cdot PE_{c1}$ $+ (2^{d-p} - 1) \cdot s_1 \cdot (PE_{c1} - 1)$	$(2^{p_1} - 1) \cdot PE_{c1}$ $+ (2^{d_1-p_1} - 1) \cdot s_1 \cdot (PE_{c1} - 1)$
$Link_{c1}$	$(2^{d-p} - 1) \cdot PE_{r1}$	$(2^{d_1-p_1} - 1) \cdot PE_{r1}$
$Block_r$	$q + s_2$	$2^{p_2 \cdot d_2} + s_2$
$Block_c$	C/q $(C = 2^{2^{h-1} \cdot (d-2) + h+1-d})$	$2^{(h-1-p_2) \cdot d_2}$
$Link_{r2}$	$(h - 1) \cdot Block_c$	$(2^{p_2 \cdot d_2} - 1) \cdot ((h - 2)/2^{p_1} + 1) \cdot Block_c^\dagger$ $+ (2^{(h-1-p_2) \cdot d_2} - 1) \cdot s_1 \cdot ((h - 2)/2^{p_1} \cdot Block_c^\dagger$ $+ (2^{(h-1-p_2) \cdot d_2} - 1) \cdot s_1 \cdot 2 \cdot (Block_c - 1)$ $+ (2^{(h-1-p_2) \cdot d_2} - 1) \cdot s_2 \cdot 2^{p_1} \cdot (Block_c - 1)^{\dagger\dagger}$
$Link_{c2}$	$(h - 1) \cdot (q - 1)$ $+ PE_{c1} \cdot PE_{r1} \cdot Block_r$	$(2^{(h-1-p_2) \cdot d_2} - 1) \cdot (2^{p_1} + s_{p_1}) \cdot Block_r^{\dagger\dagger}$

† $(h - 2)/2^{p_1}$ は整数の商。

†† $(h - 1) < 2^{p_1}$ のときは、 2^{p_1} は $h - 1$ となる。

表 4.2: 各構成法の P E 配列

	システムのサイズ	HC,HN,HHC				CCC			
		2^4	2^8	2^{12}	2^{16}	2^4	2^8	2^{12}	2^{16}
第1階層	1行のPE数 (スペアPE数)	3 (1)	5 (1)	5 (1)	9 (1)	3 (1)	5 (1)	10 (2)	18 (2)
	1列のPE数	2	2	4	4	2	2	2	2
第2階層	1行のブロック数 (スペアブロック数)	3 (1)	9 (1)	17 (1)	65 (1)	3 (1)	9 (1)	17 (1)	65 (1)
	1列のブロック数	2	4	16	32	2	4	16	32

表 4.3: 各ネットワークのレイアウト面積

		HC	CCC	HN	HHC	
(PE数) システム のサイズ 2^4	次数	4	3	3	4	
	面積 (cm^2)	$A_{com} = 0.1$	0.5491	0.4815	0.4870	0.5192
		$A_{com} = 0$	0.3982	0.3718	0.3767	0.3728
	面積の差 (cm^2)		0.1509	0.1773	0.1103	0.1464
(PE数) システム のサイズ 2^8	次数	8	3	4	8	
	面積 (cm^2)	$A_{com} = 0.1$	8.965	5.012	5.309	6.971
		$A_{com} = 0$	5.477	3.892	3.828	3.967
	面積の差 (cm^2)		3.488	1.12	1.481	3.004
(PE数) システム のサイズ 2^{12}	次数	12	3	5	12	
	面積 (cm^2)	$A_{com} = 0.1$	340.9	84.80	88.73	119.4
		$A_{com} = 0$	218.8	66.78	60.25	70.62
	面積の差 (cm^2)		122.1	18.02	28.48	48.78
(PE数) システム のサイズ 2^{16}	次数	16	3	6	16	
	面積 (cm^2)	$A_{com} = 0.1$	17870	1379	1330	2524
		$A_{com} = 0$	12810	1105	853.8	1115
	面積の差 (cm^2)		5060	274	476.2	1409

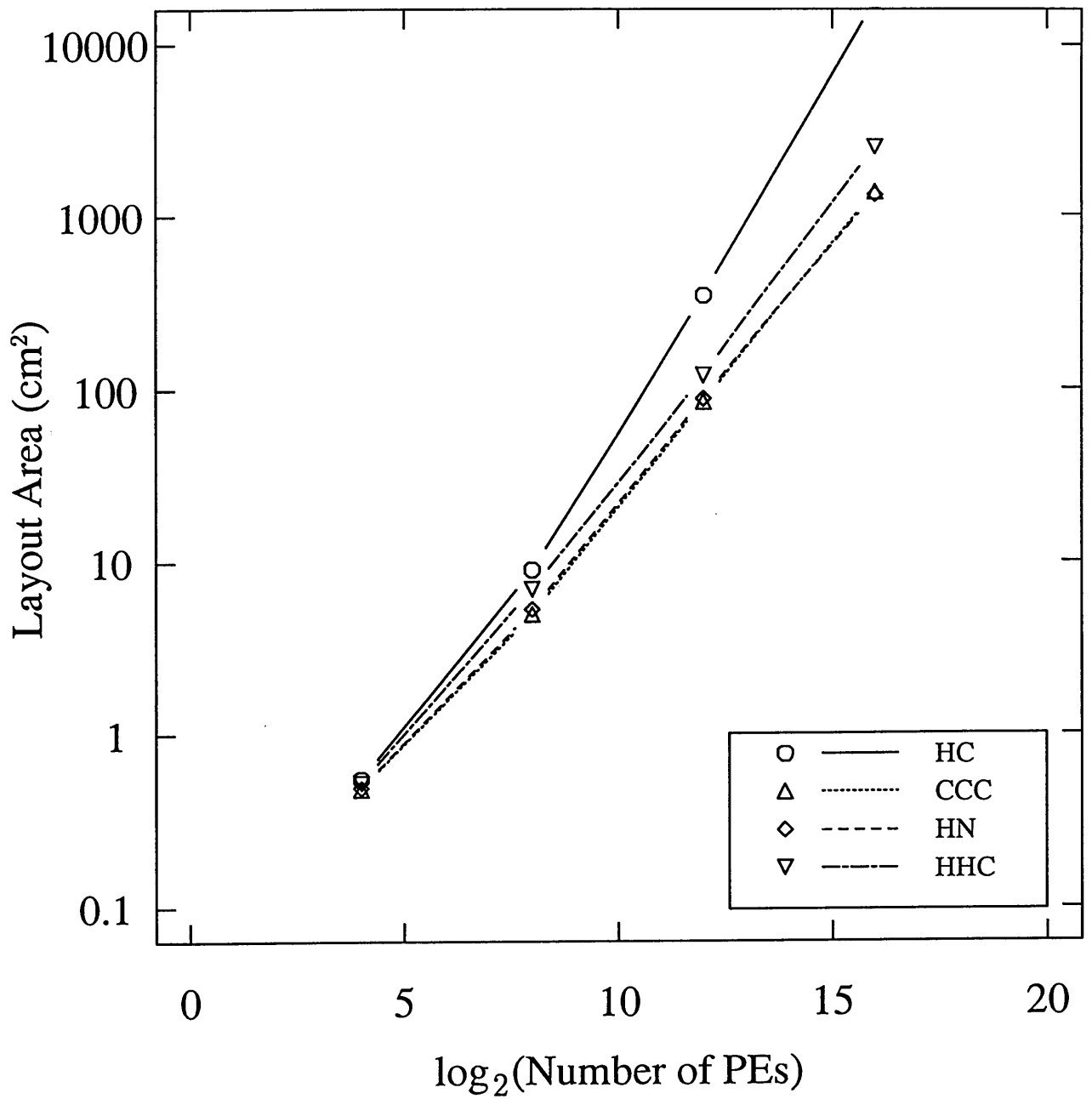


図 4.1: 通信制御回路を考慮した場合のレイアウト面積の比較
 ($W_{link} = 4 \cdot 2.0 \mu\text{m}$, $A_{com} = 0.1 \text{mm}^2$, $A_{other} = 1 \text{mm}^2$)

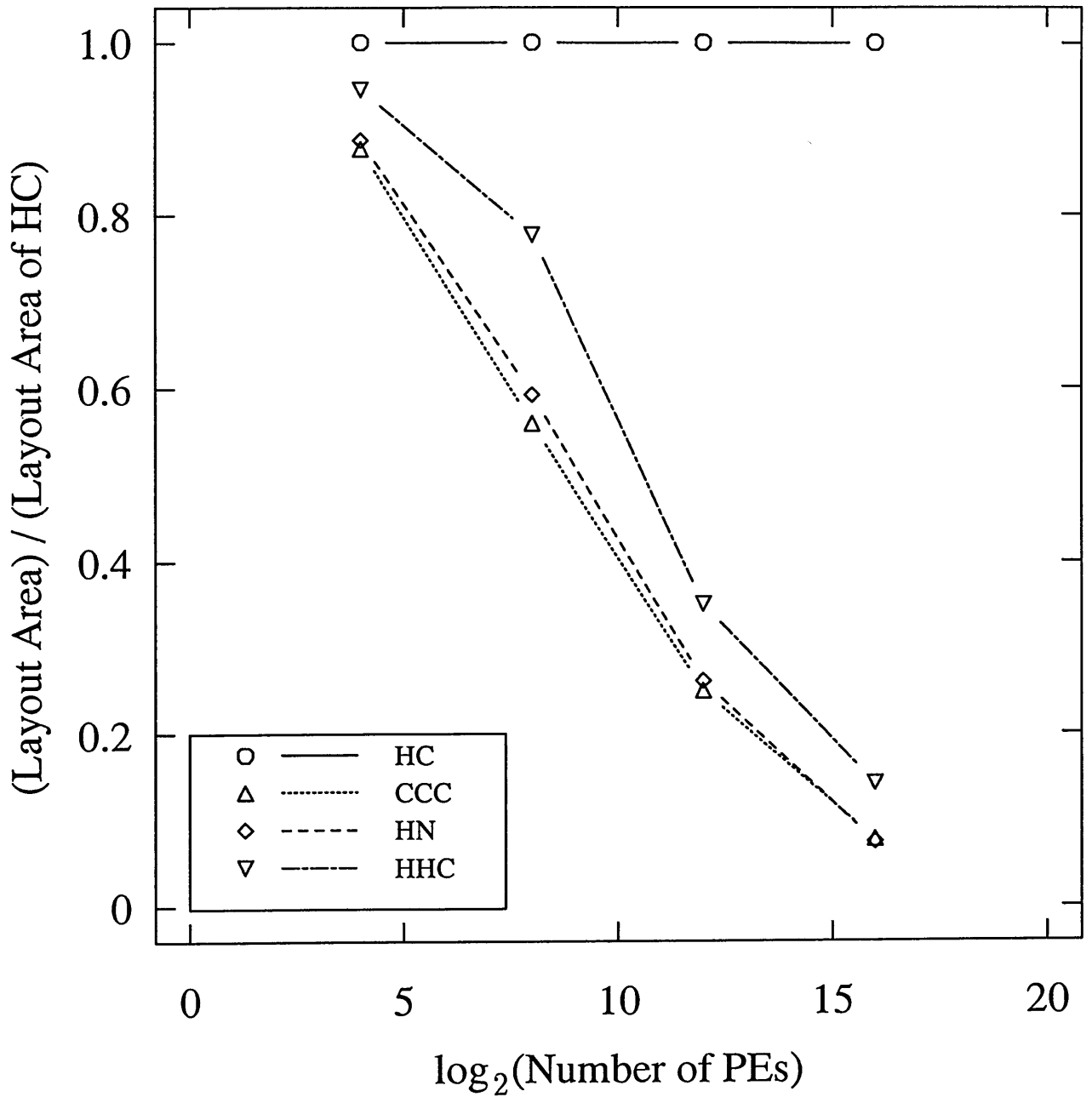


図 4.2: HCのレイアウト面積に対するレイアウト面積の比較
 ($W_{link} = 4.2.0\mu\text{m}$, $A_{com} = 0.1\text{mm}^2$, $A_{other} = 1\text{mm}^2$)

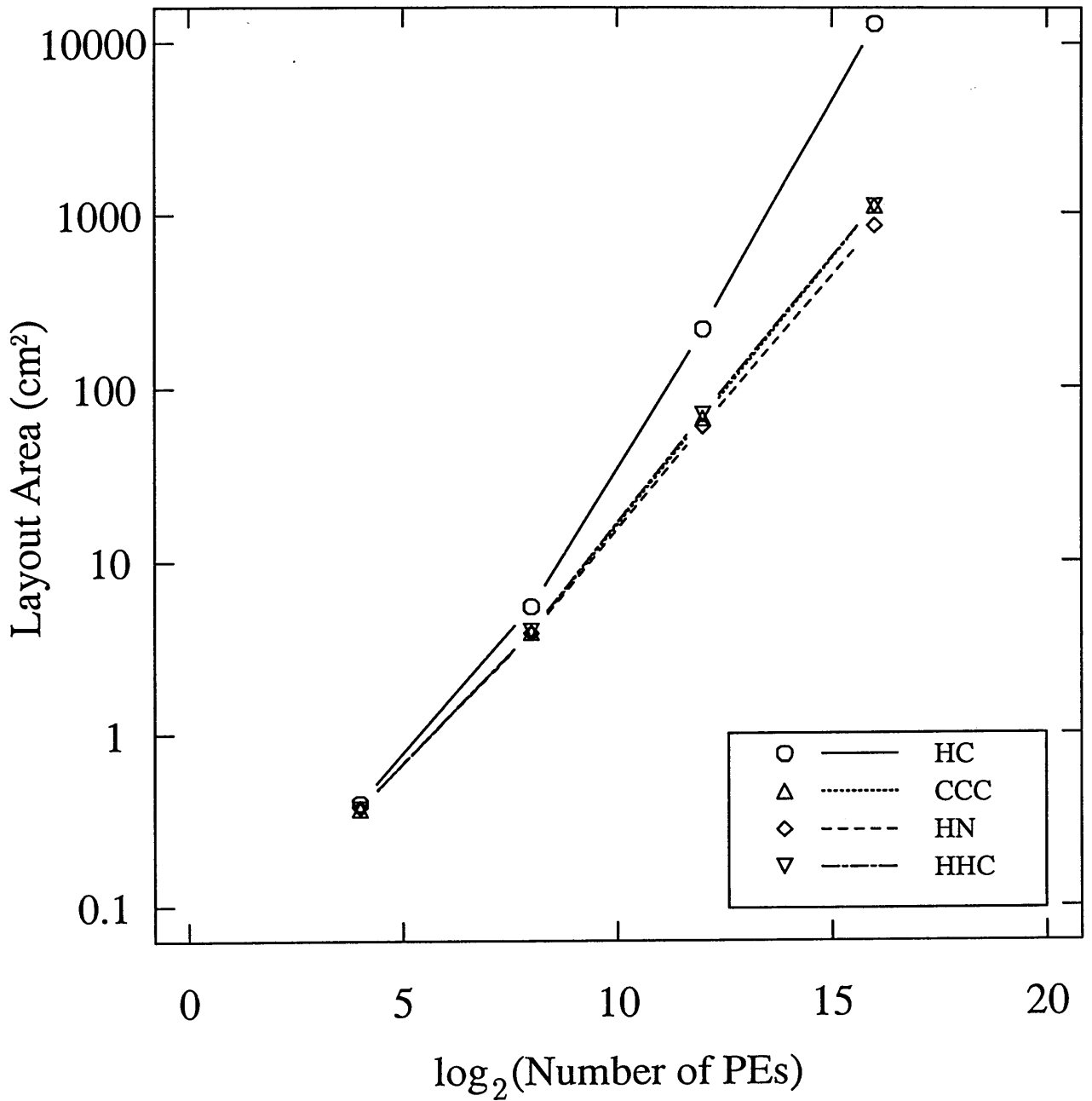


図 4.3: 通信制御回路を考慮しない場合のレイアウト面積の比較
 ($W_{link} = 4.2.0\mu\text{m}$, $A_{com} = 0\text{mm}^2$, $A_{other} = 1\text{mm}^2$)

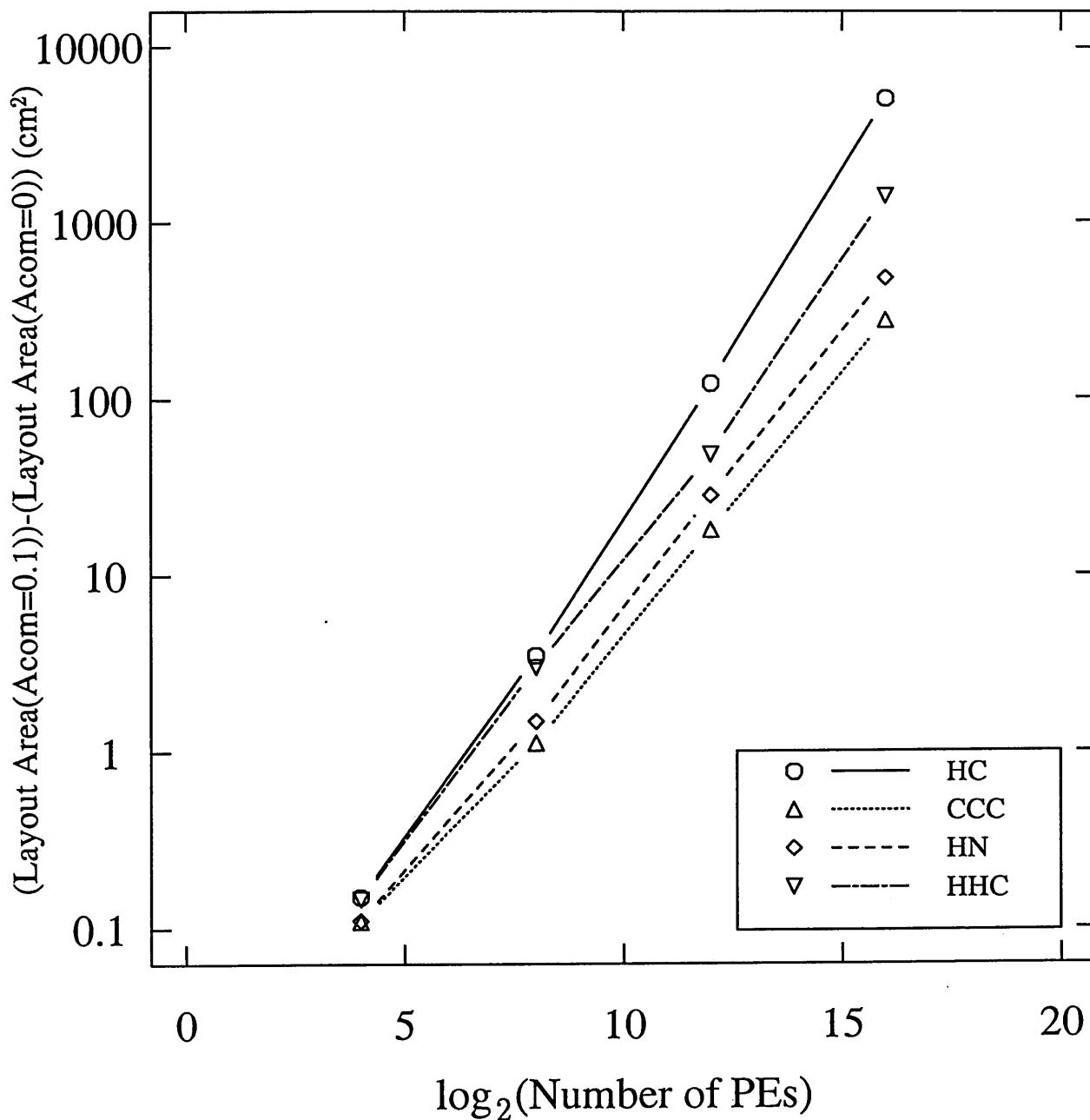


図 4.4: 通信制御回路を考慮した場合としない場合のレイアウト面積の差
 ($W_{link} = 4.2.0\mu\text{m}$, $A_{other} = 1\text{mm}^2$)

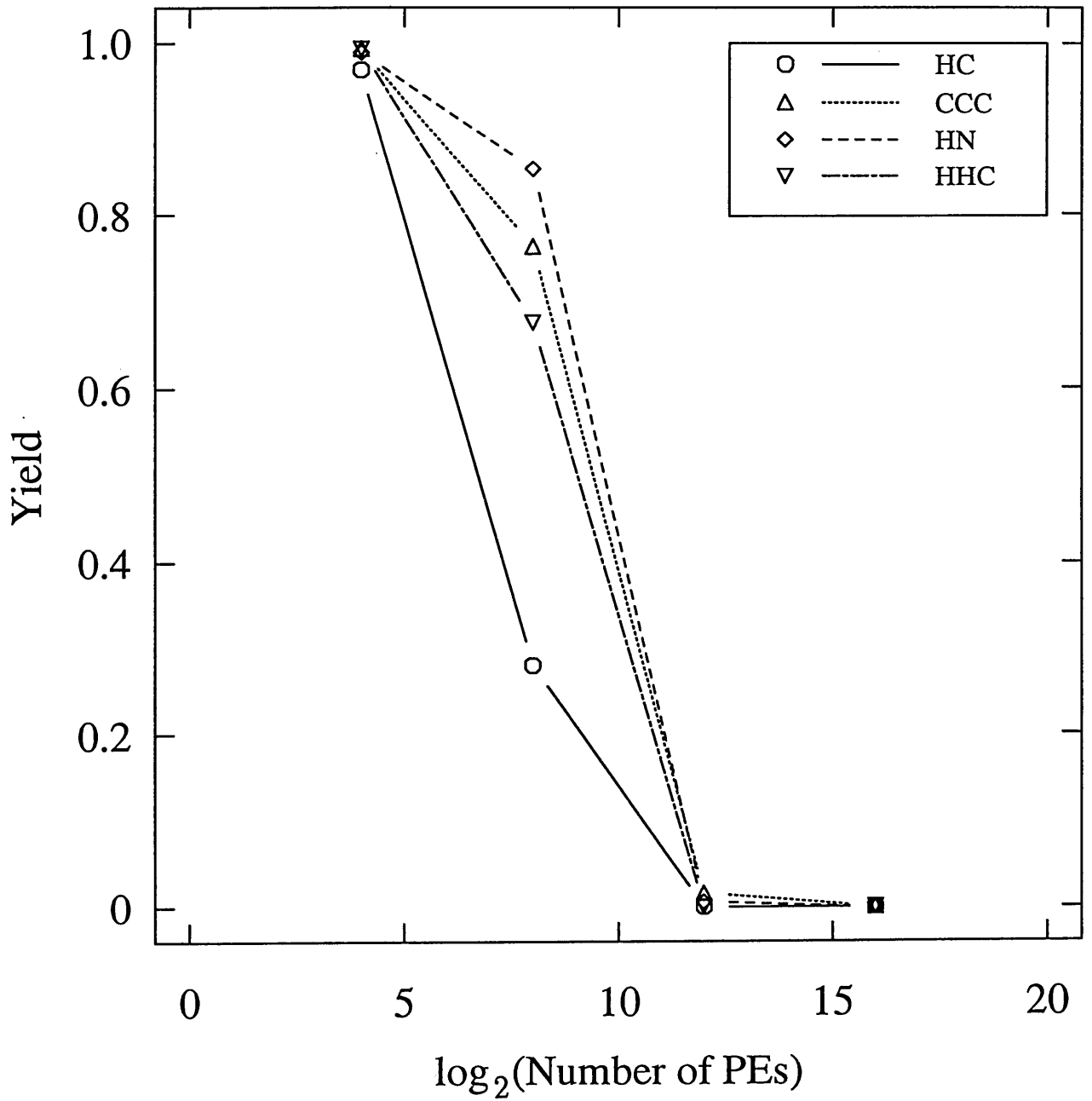


図 4.5: 通信制御回路を考慮した場合の歩留まりの比較
 ($W_{link} = 4 \cdot 2.0 \mu\text{m}$, $A_{com} = 0.1 \text{mm}^2$, $\alpha = 2.0$, $D_{pe} = 2.0 / \text{cm}^2$, $D_{link} = 1.0 / \text{cm}^2$)

第 5 章

結論

5.1 本研究のまとめ

本研究の目的は、超並列を目指すハイパキューブ結合型のマルチプロセッサシステムの W S I 構成法を検討することであった。そのために、まず第 2 章で超並列を目指すシステムのネットワークトポロジについての考察を行った。次に、第 3 章でそれらのハイパキューブ結合型のマルチプロセッサシステムに関する階層型の W S I 構成法を与え、第 4 章でその評価を行った。

ハイパキューブ結合は、汎用の並列システムのプロセッサ間接続方式として注目されている。しかし、超並列システムを目指す場合、各プロセッサ当たりのリンク接続数は増し、総リンク数も増加して実現が困難となってくる。そこで、ハイパキューブのリンク接続数と総リンク数を減らしたネットワーク構造であるハイパキューブ結合型のマルチプロセッサシステムに着目して、その性能について検討した。また、ハイパキューブを階層的に結合させたネットワーク構造を新たに考案してその性能に関して、他のハイパキューブ結合型のネットワークと比較を行った。その結果、階層型ハイパキューブはハイパネットとほぼ同じ性能を持つが、システムを構成するプロセッサ数の自由度と、P E 間の通信を行なうルーティング法によりハイパネットよりも優れていることがわかった。

これらのハイパキューブ結合型のマルチプロセッサシステムの階層型の W S I 構成法は、ネットワーク構造を 2 つの階層に分けて、それぞれの階層に冗長部分を設けて欠陥の救済を行う方法であった。階層構成にすることにより、下位の階層の冗長構成のない部分に生じた欠陥は上位の階層の冗長構成により救済することが可能となる。

さらに、W S I 構成法の評価として、レイアウト面積と歩留まりの計算法を示し、プロセッサやリンクの大きさを仮定して試算を行った。その結果、巡回型ハイパキューブ、ハイパネット、および階層型ハイパキューブはハイパキューブに比べて、レイアウト面積を大幅に減らすことができ、歩留まりを向上させることができることが分かった。

5.2 今後の課題

VLSI/WSI回路の製造時に欠陥の発生することを完全に無くすことは、将来においても、ほぼ不可能と考えられる。したがって、冗長回路構成法に関する研究は大規模なVLSI/WSI回路を製造するために不可欠なものと考えられる。

本研究の今後の課題としては、以下のものが挙げられる。

- 階層型ハイパキューブでの並列アルゴリズムの展開法の検討。
- 本研究で提案したWSI構成法の改善。
- 各ネットワークのWSI構成における、リンクの切り換えスイッチ数の比較、およびシステムの通信遅延に係る最大プロセッサ間距離の比較。
- 下位の階層において救済可能な欠陥についての考察。並びに冗長数の変化による歩留まりへの影響についての考察。
- リンクおよびスイッチの具体的な構造など、さらに細かい構成法の検討。

謝 辞

本研究を進めるにあたり、御指導、御助言を賜りました本学工学部情報工学科阿曾弘具教授に深く感謝致します。また、北陸先端科学技術大学院大学情報科学研究科堀口進教授には、研究のテーマをはじめ研究の全般にわたって様々な御助言、御指導を賜りました。さらに、多くの研究発表の機会を与えて下さいました。心より感謝致します。本論文をまとめるにあたり、本学工学部情報工学科西関隆夫教授、亀山充隆教授には適切な御意見を賜りました。深く感謝致します。本学大学院の大町真一郎氏、成富敬氏、および沼田一成氏には熱心な御討論をして頂きました。さらに、大町真一郎氏、沼田一成氏、並びに本学情報処理教育センター阿部亨助手には計算機環境を整えて頂きました。感謝致します。本学工学部情報工学科丸岡(阿曾)研究室の修士2年池田啓明氏、越後和徳氏、張新学氏、および中井満氏には日頃お世話になりました。ありがとうございました。最後にゼミや日常生活において貴重な意見や助言を下さいました本学工学部情報工学科丸岡(阿曾)研究室の皆様、並びに卒業生の方々にもお礼を申し上げます。

参考文献

- [1] 堀口 進: “ウェハ規模超密度集積回路について”, *HYBRID*, 6, 1 (1990).
- [2] S.Horiguchi, “Wafer Scale Integration”, *Proc. 6th Int'l Microelectronics Conf. SHM*, pp.51-58 (1990).
- [3] 南谷 崇: “フォールトトレランス技術の最近の動向 [II]”, *信学誌*, 73, 10, pp.1095-1102 (1990-10).
- [4] M.Chean and J.Forters, “A Taxonomy of Reconfiguration Techniques for Fault-Tolerant Processor Arrays”, *Computer*, 23, 1, pp.55-69 (Jan. 1990).
- [5] J.H.Kim and S.M.Reddy, “On the Design of Fault-Tolerant Two-Dimensional Systolic Arrays for Yield Enhancement”, *IEEE Trans. Comput.*, 38, 4, pp.515-525 (Apr. 1989).
- [6] 伊藤 秀男, 鈴木 信行: “ウェハスケール LSI 上でのハイパキューブの構成法”, *信学論 (D-I)*, J73-D-I, 3, pp.314-323 (1990-03).
- [7] 伊藤 秀男, 細谷 英一, 管谷: “基本チャネルを用いる W S I ハイパキューブのアレー構造”, *信学会 ウェハスケール集積システム研究会*, WSI-91-9 (1992-02).
- [8] 細谷 英一, 伊藤 秀男: “シフトチャネルを用いた W S I ハイパキューブの構成法”, *信学会 ウェハスケール集積システム研究会*, WSI-92-5 (1992-08).
- [9] F.P.Preparata and J.Vuillemin, “The Cube-Connected Cycles: A Versatile Network for Parallel Computation”, *Commun. ACM*, 24, 5, pp.300-309 (May 1981).
- [10] K.Hwang and J.Ghosh, “Hypernet: A communication-efficient architecture for constructing massively parallel computers”, *IEEE Trans. Comput.*, 36, 12, pp.1450-1466 (Dec. 1987).
- [11] 山下 公一 他: “ウェハ・スケール L S I の可能性と限界”, *日経エレクトロニクス*, 372, pp.141-161 (1987).

- [12] C.L.Seitz, "The cosmic cube", *Commun. ACM*, **28**, 1, pp.22-33 (Jan. 1985).
- [13] D.L.Waltz, "Application of the connection machine", *Computer*, **20**, 1, pp.85-97 (Jan. 1987).
- [14] L.D.Wittie, "Communication Structures for Large Networks of Microcomputers", *IEEE Trans. Comput.*, **30**, 4, pp.264-273 (Apr. 1981).
- [15] S.P. Dandamudi and D.L. Eager, "Hierarchical Interconnection Networks for Multi-computer Systems", *IEEE Trans. Comput.*, **39**, 6, pp.786-797 (Jun. 1990).
- [16] S.Y.Kuo and W.K.Fuchs, "Reconfigurable Cube-Connected Cycles Architectures", *Jour. Parallel and Distributed Computing*, **9** pp.1-10 (1990).
- [17] N.F.Tzeng, "A Reconfigurable Cube-Connected Cycles Architecture for Wafer Scale Integration", *Proc. Int'l Conf. Wafer Scale Integration*, pp.33-39 (Jan. 1991).
- [18] 福田 大, 堀口 進: "巡回型ハイパキューブ結合のWSI構成方式", 信学会 ウェハスケール集積システム研究会, **WSI-91-10** (1992-02).
- [19] M.Wang, M.Cutler and S.Y.H.Su, "Reconfiguration of VLSI/WSI mesh array processors with two-level redundancy", *IEEE Trans. Comput.*, **38**, 4, pp.547-553 (Apr. 1989).
- [20] 津田 伸生: "欠陥救済型切換え回路を用いた2次元直交アレーの階層化冗長構成法", 信学論 (D-I), **J75-D-I**, 4, pp.251-260 (1992-04).
- [21] R.Suaya and G.Birtwistle, "VLSI and Parallel Computation", *Morgan Kaufmann*, pp.199-204.
- [22] C.H.Stapper, F.M.Armstrong and K.Saji, "Integrated circuit yield statics", *Proc. IEEE*, **71**, 4, pp.453-470 (Apr. 1983).
- [23] I.Koren, Z.Koren and D.K.Pradhan, "Designing Interconnection buses in VLSI and WSI for maximum delay", *IEEE. J. Solid-State Circuits*, **23**, 3, pp.859-866 (Mar. 1988).
- [24] 竹内 啓, 藤野 和建: "2項分布とポアソン分布", 東京大学出版会, pp.51-53 (1981).

研究業績一覧

- 論文

- ・ 巡回型ハイパキューブ結合のWSI構成方式
福田 大、堀口 進
電子情報通信学会論文誌 (D-I)(1993-02 掲載予定)

- 学会・研究会

1. サイクリックキューブ結合型マルチプロセッサシステムのWSIアーキテクチャ
福田 大、堀口 進
平成3年度第1回情報処理学会東北支部研究会 3-1-4 (1991-05)
2. サイクリックキューブ結合型マルチプロセッサシステムのWSIアーキテクチャ
福田 大、堀口 進、木村 正行
平成3年度電気関係学会東北支部連合大会 2H24 (1991-08)
3. 巡回型ハイパーキューブ結合のWSI構成方式
福田 大、堀口 進
電子情報通信学会ウェハスケール集積システム研究会 WSI-91-10 (1992-02)
4. 巡回型ハイパキューブ結合マルチプロセッサのWSI構成法
堀口 進、福田 大
電子情報通信学会ウェハスケール集積システム研究会 WSI-92-8 (1992-08)
5. de Bruijn ネットワークのWSI構成法
福田 大、堀口 進
平成4年度電気関係学会東北支部連合大会 2B6 (1992-08)
6. 階層型ハイパキューブ結合のWSI構成法
福田 大、堀口 進
情報処理学会第45回全国大会 7L-6 (1992-10)