第２章　相互結合網

２．１相互結合網の分類

２.１.１分類項目

（１）制御方式

①集中制御方式

②分散制御方式

（２）交換方式

①パケット交換方式（packet switching**）：**

手紙による通信

・蓄積交換（store and forward）

・ワームホール（worm hole）

・バーチャルカットスルー

（virtual cut through）

①回線交換方式（circuit switching）：

電話による通信

（３）トポロジ

静的網：直接結合（direct connection）網

完全網を規則的に簡略化

動的網：間接結合（Indirect Connection）網

クロスバスイッチを規則的に簡略化

２．１．２評価項目

（1）距離（distance）

（２）次数（degree）

（３）総スイッチ数／総リンク数

（４）拡張性（scalability）

（５）３次元実装の容易性

（６）耐故障性（fault tolerance）

（７）多様な網の埋込み能力（embedability）

静的網の場合：

多数の他の網の埋込みが可能である必要

並列処理の応用プログラム

論理構造に適合した特定のトポロジ上で

最も高速に実行

（８）ルーティングの容易性

２．１．３基本通信パターン

送信ノード番号Ｘの２進表示:$(a_n,...,a_2,a_1)$

受信ノード番号Ｙの２進表示：$(b_n,...,b_2,b_1)$

通信パターンの数（受信ノードにダブリなし）：$N!$

送受信ノードの対応関係：ひとつの置換

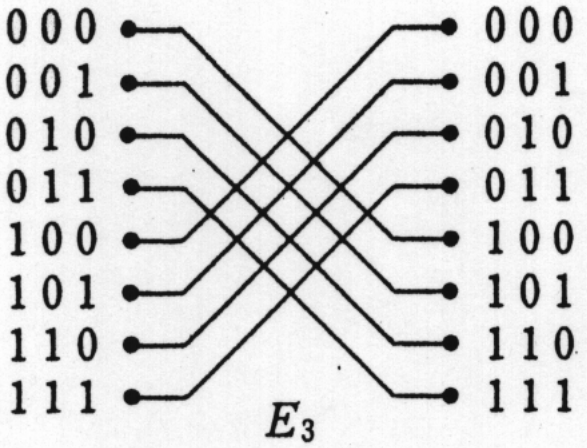
送信　　受信

（１，２，３，４）＜－＞（２，３，４，１）

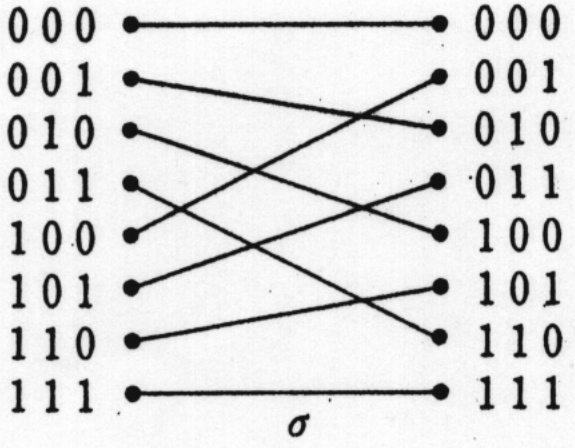
（１）エクスチェンジ置換

$$y=E_i(x)=(a_n,...,\sim a_i,...,a_1)$$

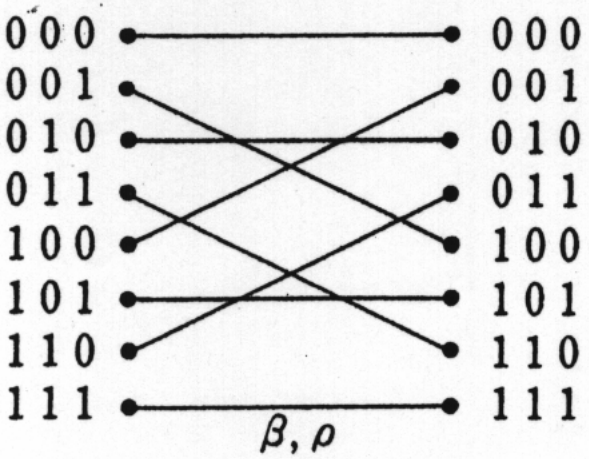
～a_iはa_iの否定を表す。



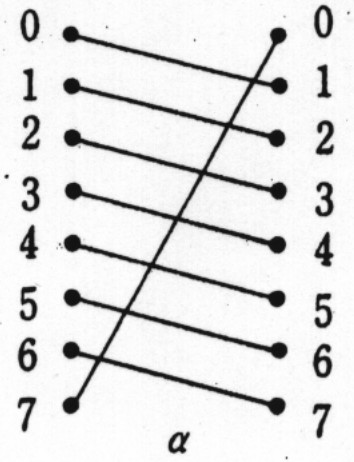
（a） エクスチェンジ置換



（b） シャフル置換



（c） バタフライ置換と
ビット逆転置換



（d） シフト置換

（２）シャフル置換

$$y=\sigma(x)=(a_{n-1},a_{n-2},\ldots,a_1,a_n)$$

ｋ－サブシャフル σ_k 、

ｋ－スーパシャフル σ^k

$$y=\sigma_k(x)=(a_n,\ldots,a_{k+1},a_{k-1},\ldots,a_1,a_k)$$

$$y=\sigma^k(x)=(a_{n-1},\ldots,a_{n-k+1},a_n,a_{n-k},\ldots,a_1)$$

（３）バタフライ置換

$$y=\beta(x)=(a_1,a_{n-1},\ldots,a_2,a_n\ )$$

ｋ－サブバタフライ　β_k、

ｋ－スーパバタフライ　β^k

$$y=\beta_k(x)=(a_n,\ldots,a_{k+1},a_1,a_{k-1},\ldots,a_2,a_k)$$

$$y=\beta^k(x)=(a_{n-k+1},a_{n-1},\ldots,a_{n-k+2},a_n,a_{n-k},.,a_1)$$

（４）ビット逆転置換

$$y=\rho(x)=(a_1,a_2,\ldots,a_n)$$

ｋ－サブビット逆転置換　ρ_k、

ｋ－スーパビット逆転置換　ρ^k

$$y=\rho_k(x)=(a_n,\ldots,a_{k+1},a_1,a_2,\ldots,a_{k-1},a_k)$$

$$y=\rho^k(x)=(a_{n-k+1},a_{n-k+2},\ldots,a_{n-1},a_n,a_{n-k},.,a_1)$$

（５）シフト置換

$$y=\alpha(x)=|x+1|2^n$$

ｋ－サブシフト置換α_k 、

ｋ－スーパシフト置換 α^k

$$y=\alpha_k(x)=|x+1|2^k+\lfloor x/2^k \rfloor 2^k$$

$$y=\alpha^k(x)= |x+2^{n-k}|2^n$$

||：モジュロ計算、

$\lfloor a \rfloor$：床関数（floor function）

恒等置換 i

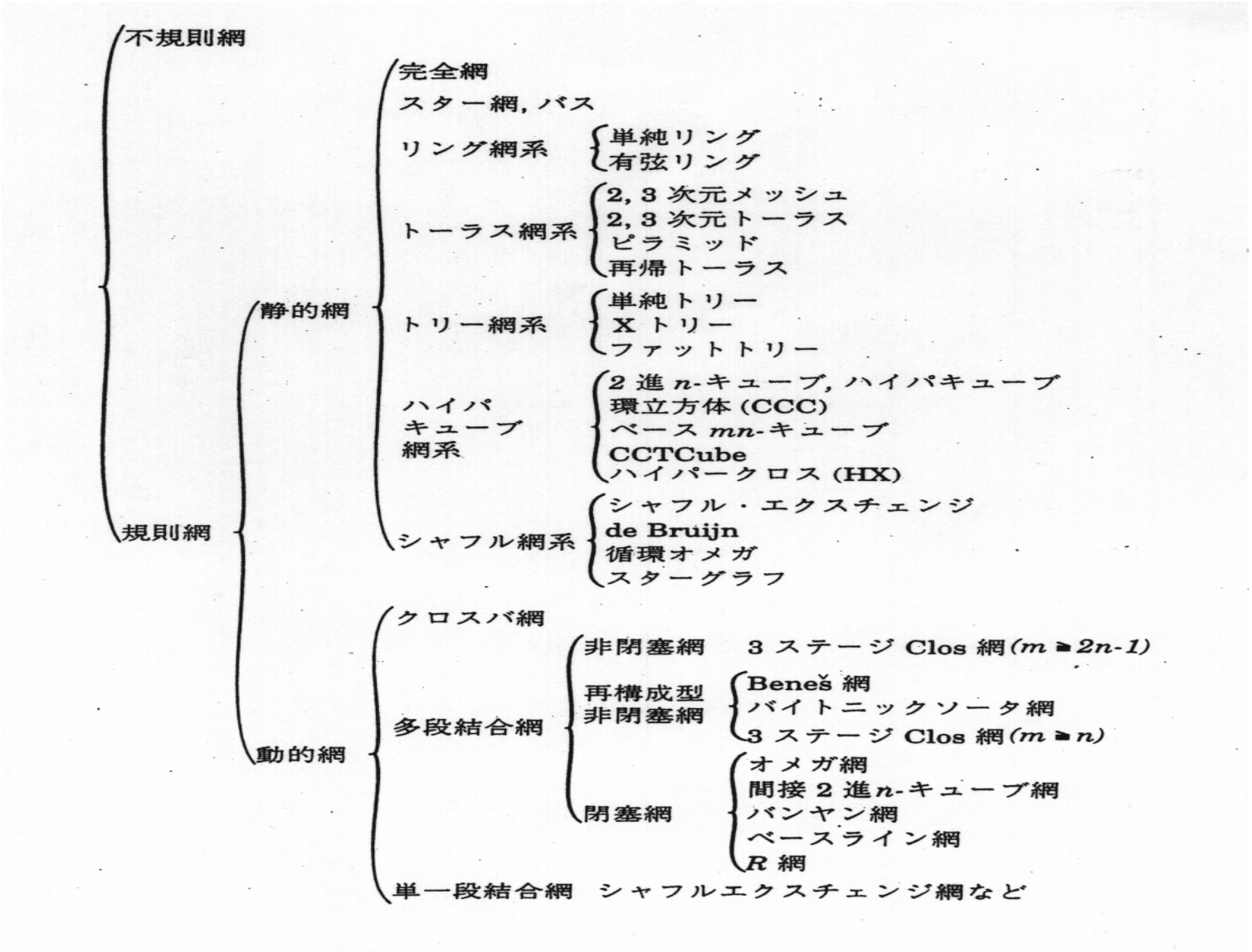
$$i(x)=x$$

逆置換π^{-1}

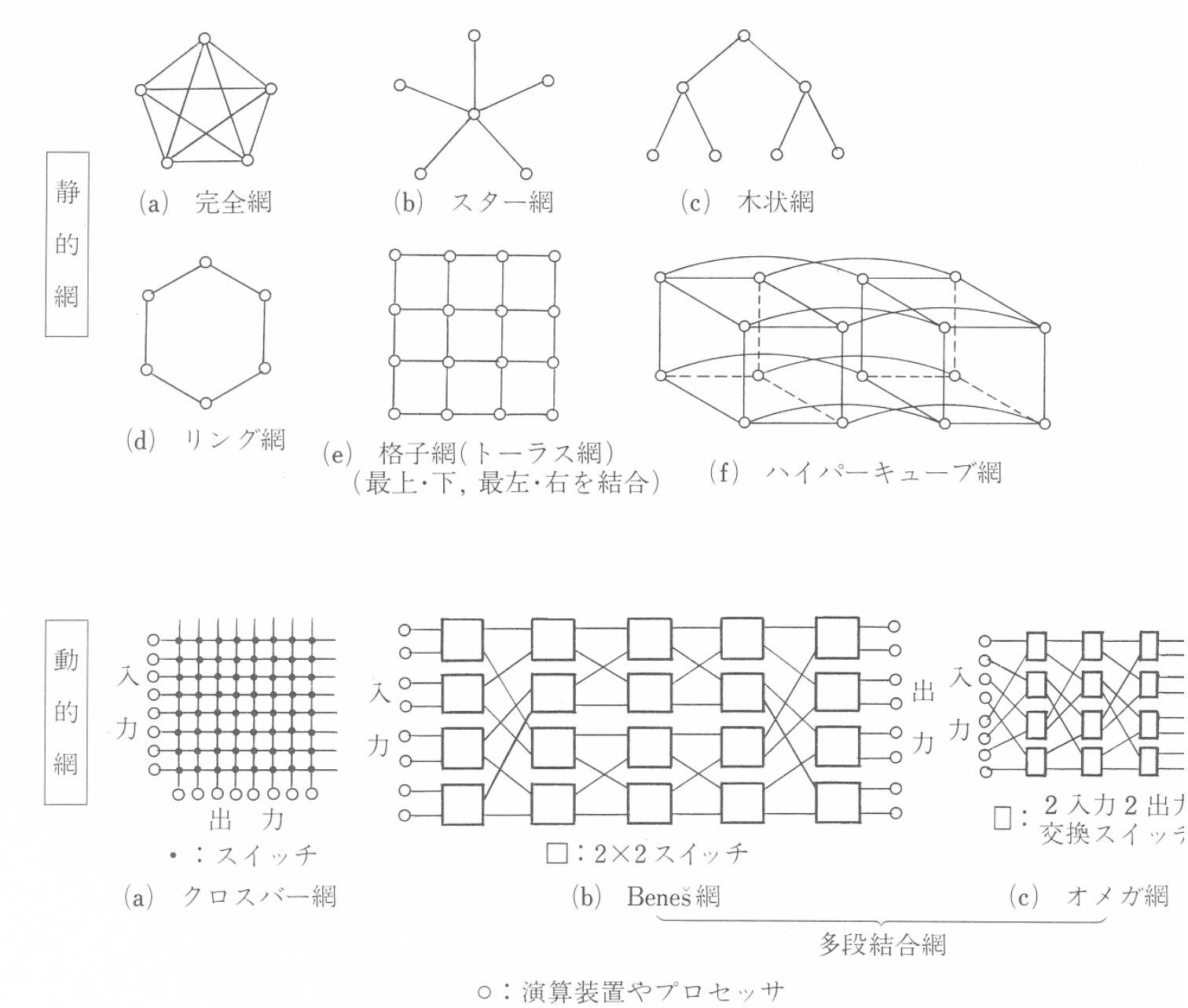
$$\pi\pi^{-1}=\pi^{-1}\pi=i$$

逆シャフル置換

$$y=\sigma^{-1}(x)=(a_1,a_n,a_{n-1},a_{n-2},\ldots,a_2)$$


不規則網
規則網
静的網
完全網
スター網, バス
リング網系
単純リング
有弦リング
トーラス網系
2, 3 次元メッシュ
2, 3 次元トーラス
ピラミッド
再帰トーラス
トリー網系
単純トリー
X トリー
ファットトリー
ハイパ
キューブ
網系
2 進 n-キューブ, ハイパキューブ
環立方体 (CCC)
ベース mn-キューブ
CCTCube
ハイパークロス (HX)
シャフル網系
シャフル・エクスチェンジ
de Bruijn
循環オメガ
スターグラフ
動的網
クロスバ網
多段結合網
非閉塞網
3 ステージ Clos 網 (m ≧ 2n-1)
再構成型
非閉塞網
Beneš 網
バイトニックソータ網
3 ステージ Clos 網 (m ≧ n)
閉塞網
オメガ網
間接 2 進 n-キューブ網
バンヤン網
ベースライン網
R 網
単一段結合網
シャフルエクスチェンジ網など

7．2相互結合網


静的網
(a) 完全網
(b) スター網
(c) 木状網
(d) リング網
(e) 格子網(トーラス網)
(最上・下，最左・右を結合)
(f) ハイパーキューブ網
動的網
入力
出力
・：スイッチ
(a) クロスバー網
入力
出力
□：2×2スイッチ
(b) Benes̆網
入力
□：2入力2出力
交換スイッチ
(c) オメガ網
多段結合網
○：演算装置やプロセッサ

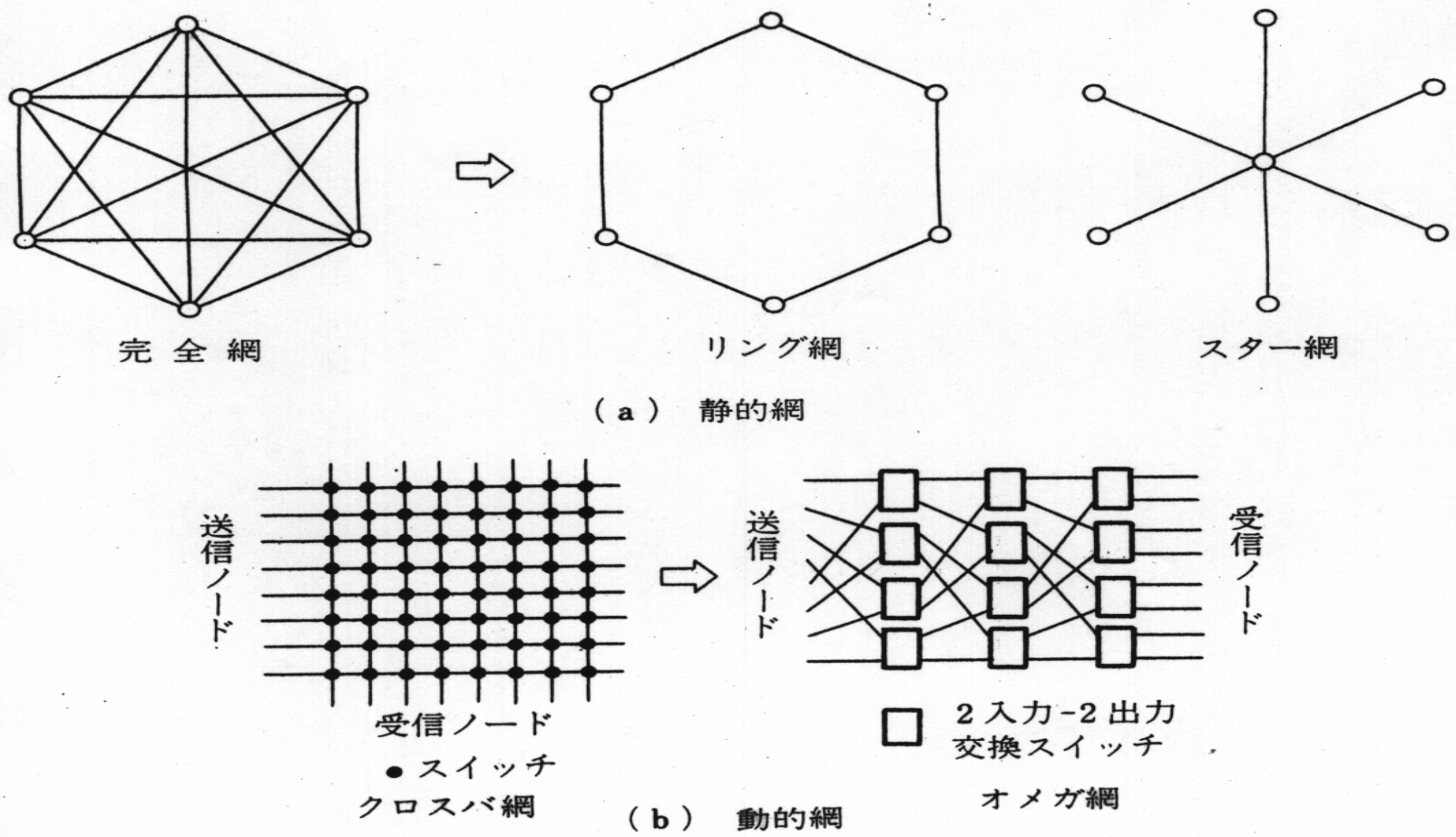

完 全 網
リング網
スター網
（a） 静的網
送信ノード
受信ノード
● スイッチ
クロスバ網
送信ノード
受信ノード
2入力-2出力
交換スイッチ
オメガ網
（b） 動的網

図 2.2　静的網と動的網

２．２静的網

２.２.１分類

（１）完全網

（２）スター網

（３）リング網系

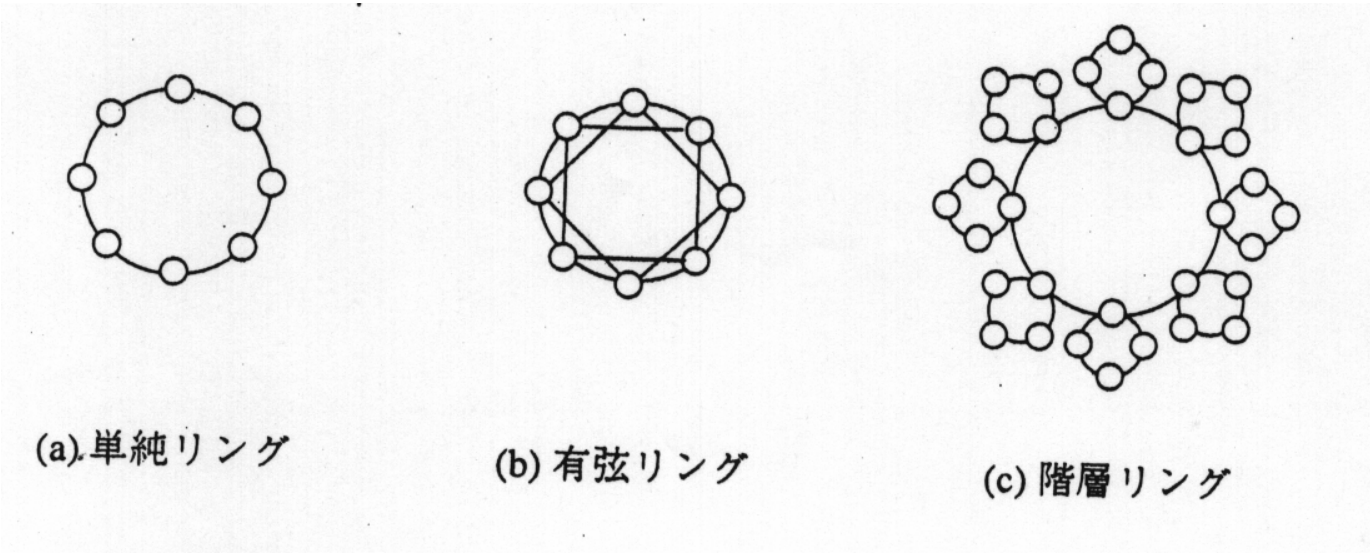
（４）トーラス網系

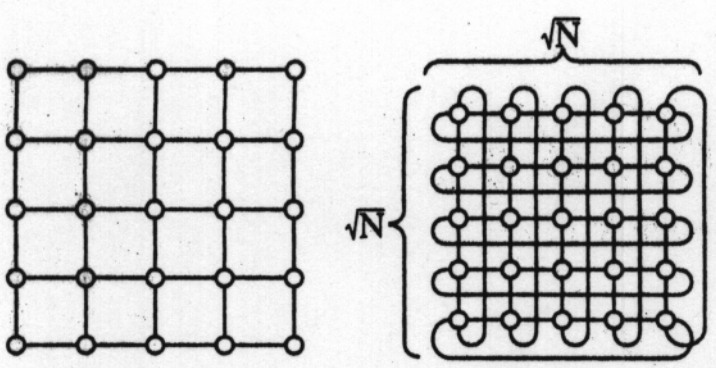
①２次元メッシュ網

②２次元トーラス網

③３次元トーラス網

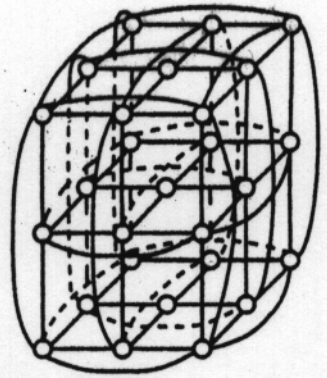
④ピラミッド網


(a) 単純リング
(b) 有弦リング
(c) 階層リング

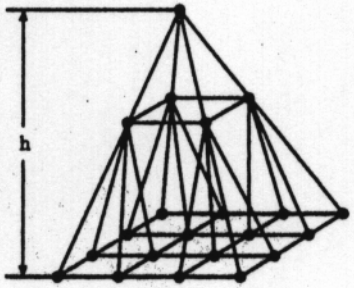


(a)2次元メッシュ

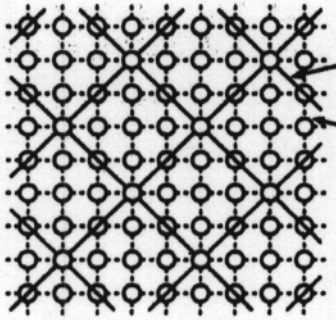
(b)2次元トーラス



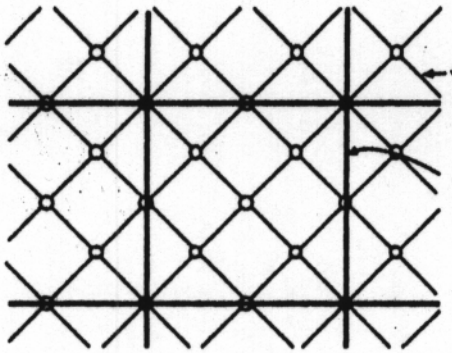
(c)3次元トーラス



(d)ピラミッド



(i)ランク1トーラス



(ii)ランク2トーラス

(e)再帰トーラス

③再帰トーラス網

（RDT、Recursive Diagonal Torus）

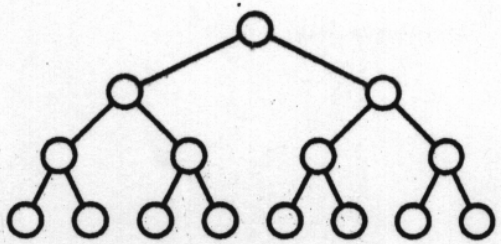
直径はN=1024で７、4096で８、16384で１０程度

（５）トリー網系

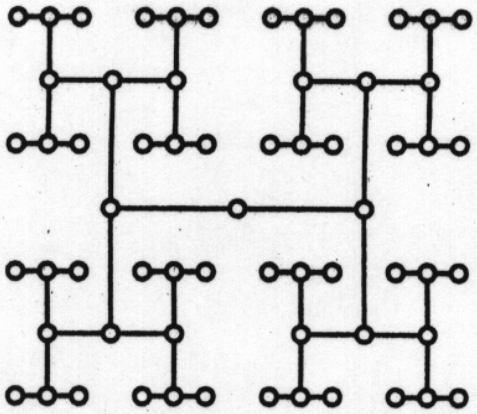
①単純トリー網

②Ｘトリー

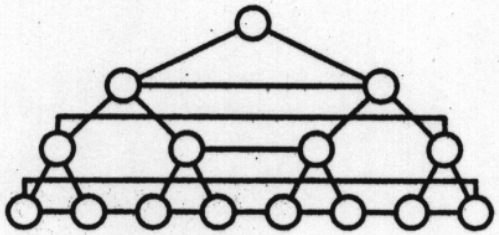
③ファットトリー（fat tree）



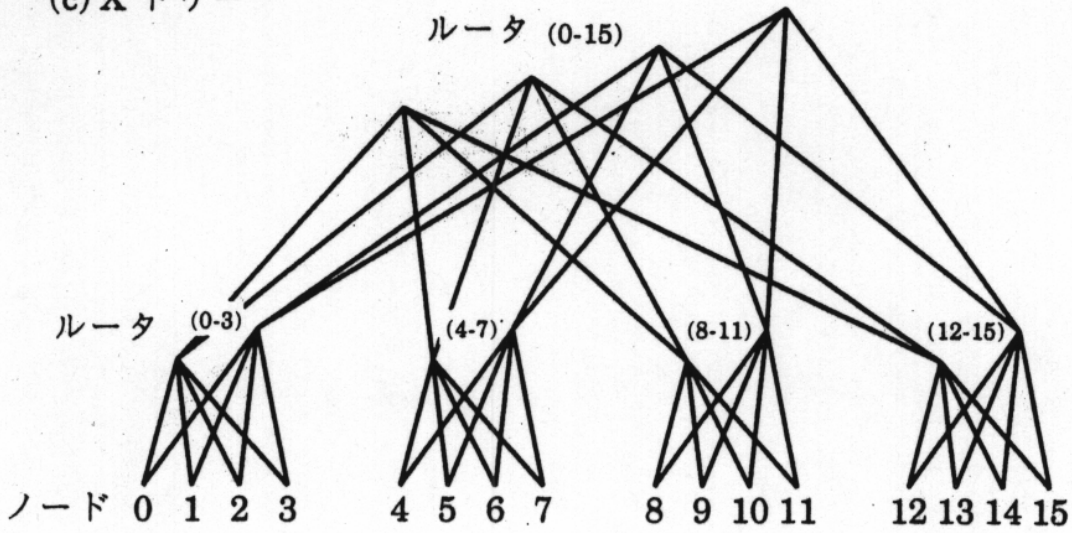
(a) 単純トリー



(b) トリーの VLSI チップへの埋込み



(c) X トリー



(d) ファットトリー (CM-5)

（６）ハイパーキューブ網系

①ハイパーキューブ網

特徴

- ルーティング容易
- 拡張可能性
- 多様な網の埋め込み
  - リング：ノード番号を反射２進で
  - トーラス：ノード番号２次元反射２進で
  - トリー：2^nの１つをダミーノード

結合

$(a_n, a_{n-1}, , , , a_2, a_1)$

$(a_n, a_{n-1}, , , , a_2, \overline{a}_1)$

$(a_n, a_{n-1}, , , , \overline{a}_2, a_1)$

$(a_n, \overline{a}_{n-1}, , , , a_2, a_1)$

$(\overline{a}_n, a_{n-1}, , , , a_2, a_1)$

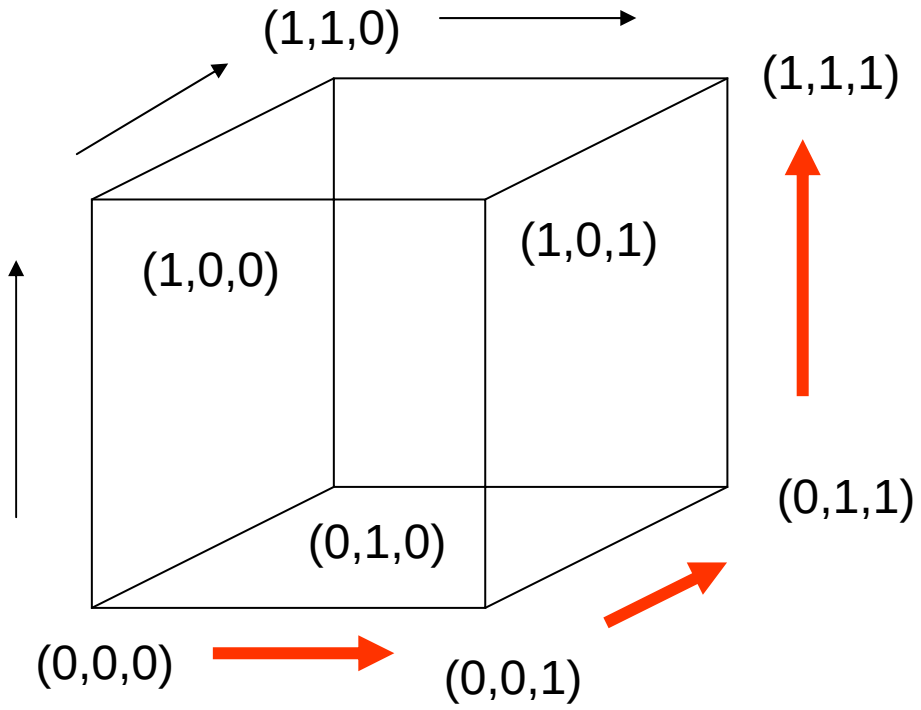
ハミング距離：1

(001)　　(110)

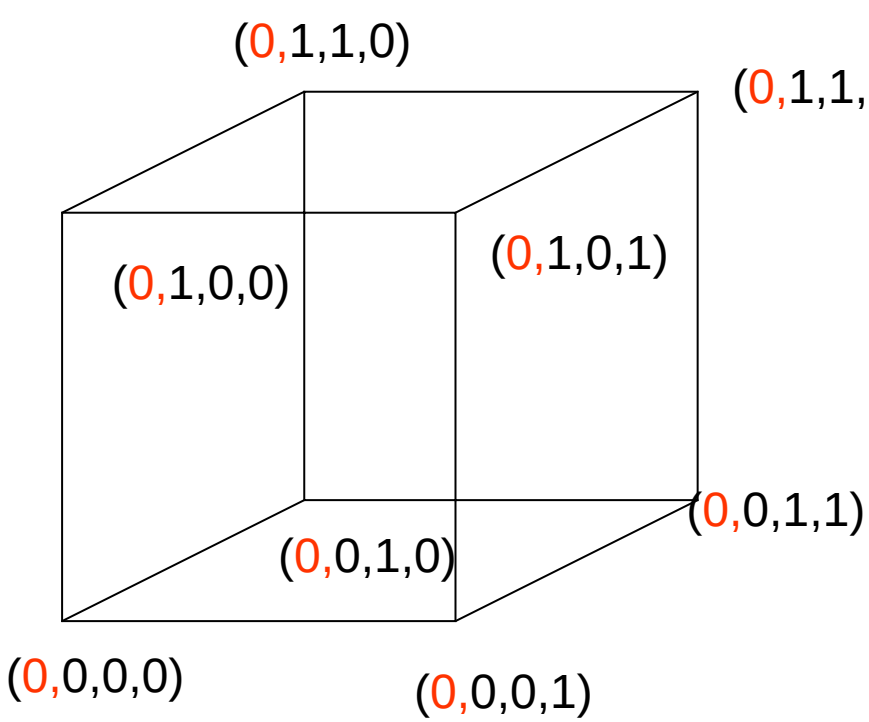
ハミング距離：3

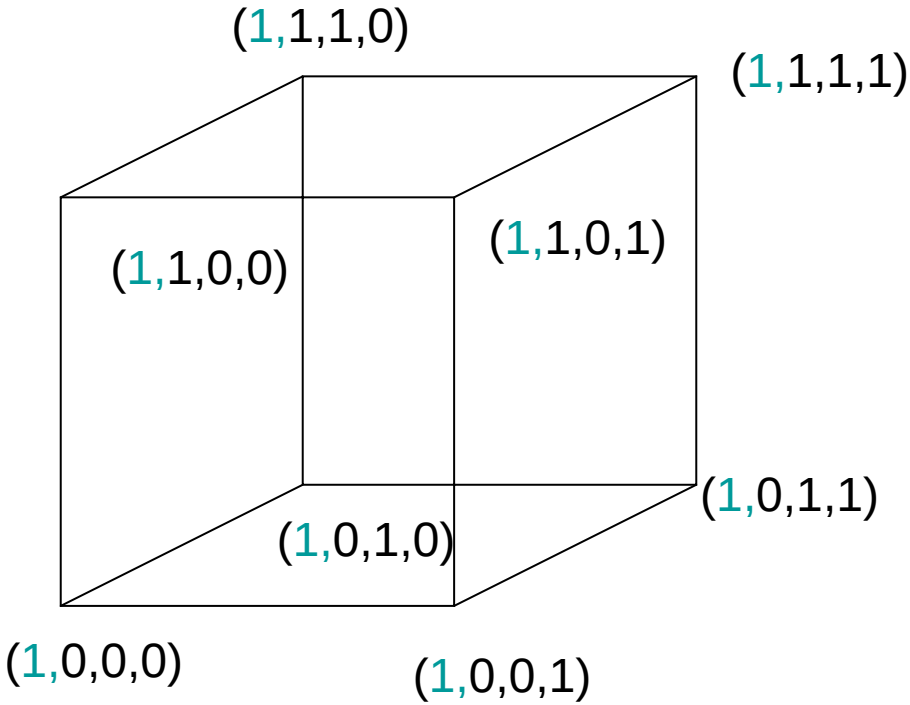
ハミング距離：各桁の比較。異なる桁の数

ルーティング
下位ビットから宛先に合わせる

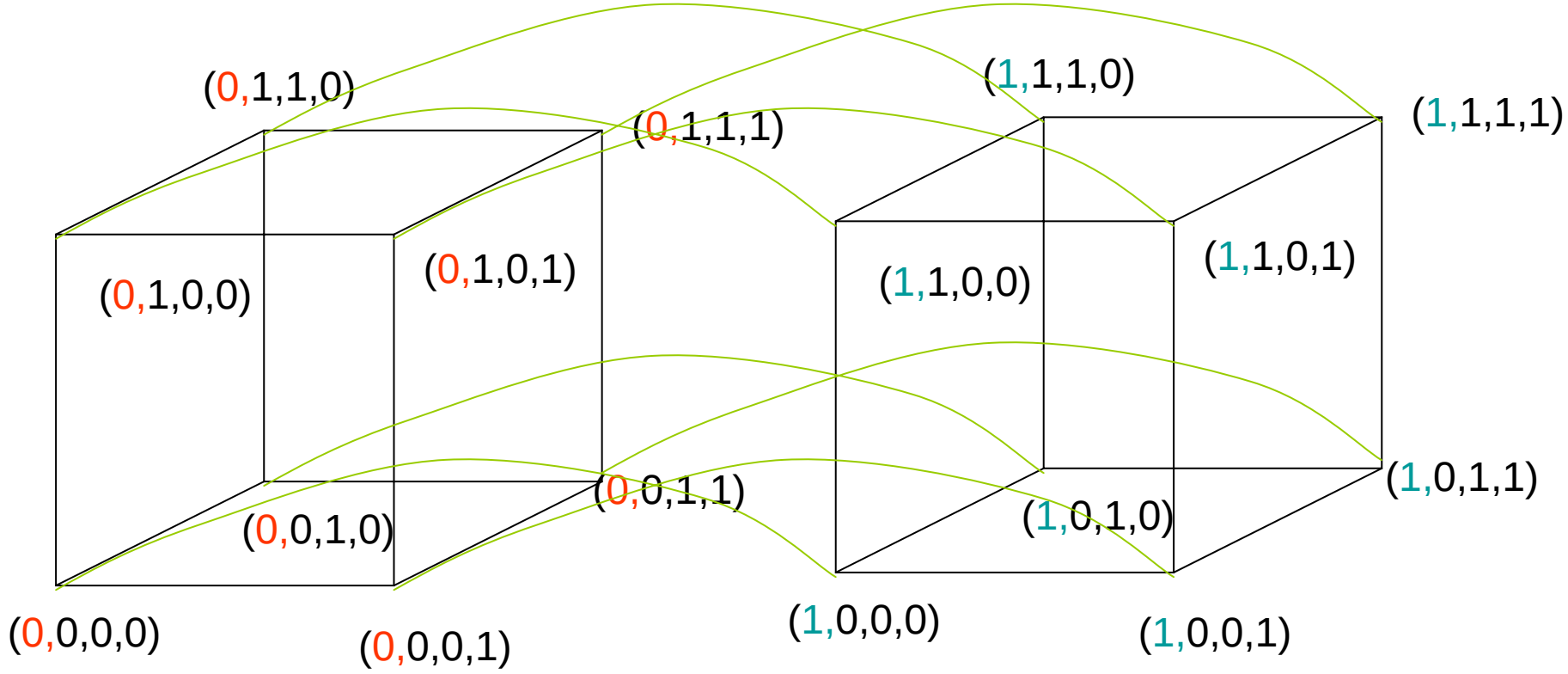


拡張は？


(0,1,1,0)
(0,1,1,1)
(0,1,0,0)
(0,1,0,1)
(0,0,1,1)
(0,0,1,0)
(0,0,0,0)
(0,0,0,1)


(1,1,1,0)
(1,1,1,1)
(1,1,0,0)
(1,1,0,1)
(1,0,1,1)
(1,0,1,0)
(1,0,0,0)
(1,0,0,1)

拡張
3キューブから4キューブへ



埋め込み能力は

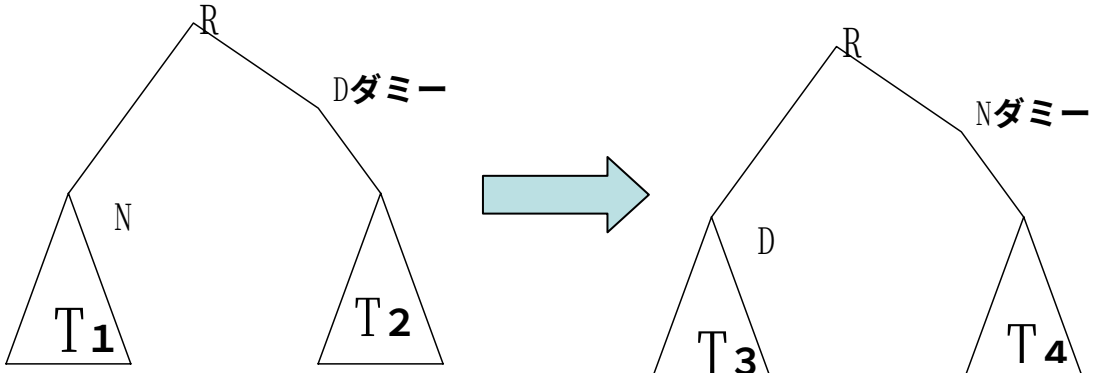
- リング
  - 1次元の反射2進符号
- トーラス
  - 2次元の反射2進符号
- トリー
  - 1つのダミーノードを許すと可能

反射2進符号
符号間のハミング距離＝1

	通常2進		反射2進		
	000		000		000
1	001	1	001		001
2	010	1	011		011
1	011	1	010		010
3	100	1	110		110
1	101	1	111		111
2	110	1	101		101
1	111	1	100		100

トリーの埋め込み

証明



n=k

$R:(a_n a_{n-1}...a_N...a_D...a_1)$

$D:(a_n a_{n-1}...a_N...\bar{a}_D...a_1)$で始まるダミー＋T2

$N:(a_n a_{n-1}...\bar{a}_N...a_D...a_1)$で始まるT1

$(a_n a_{n-1}...a_N...\bar{a}_D...a_1)$で始まるT3

$(a_n a_{n-1}...\bar{a}_N...a_D...a_1)$で始まるダミー＋T4

が存在する。

$R:(a_n a_{n-1}...a_N...a_D...a_1)$

$R':(a_n a_{n-1}...a_D...a_N...a_1)$

RとR'は1対１対応

$(a_n a_{n-1}...\bar{a}_N...a_D...a_1),(a_n a_{n-1}...a_N...\bar{a}_D...a_1)$から始まる

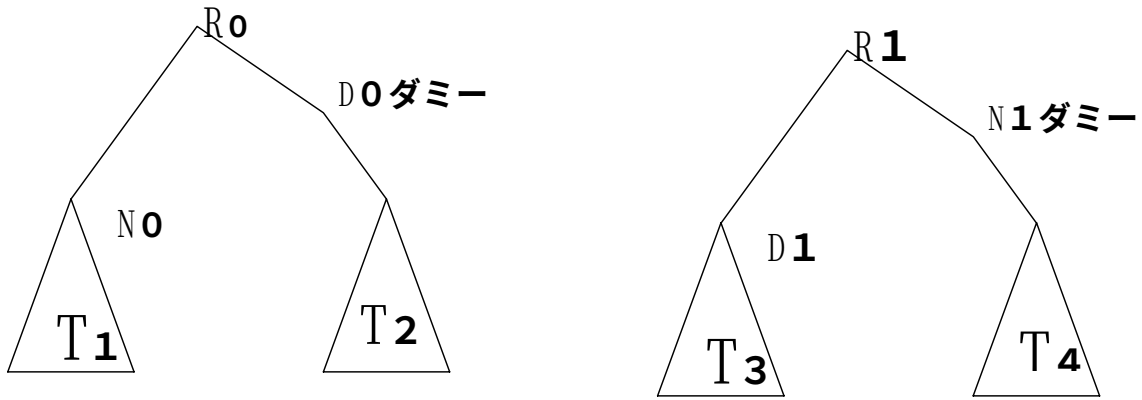
トリーに重なりがなければ

$(a_n a_{n-1}...\bar{a}_D...a_N...a_1),(a_n a_{n-1}...a_D...\bar{a}_N...a_1)$から始まる

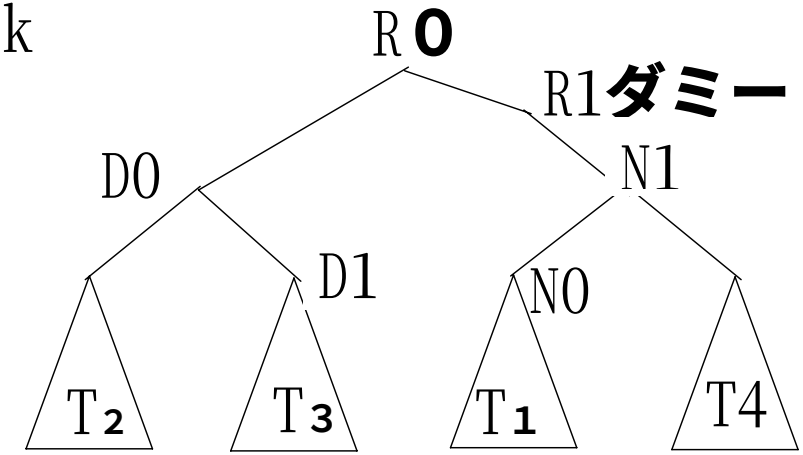
トリーに重なりがない。

トリーの埋め込み

証明



n=k



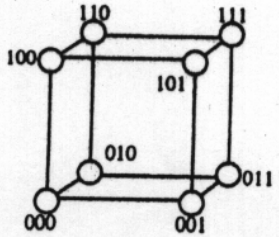
n=k+1

②ベースｍｎ－キューブ

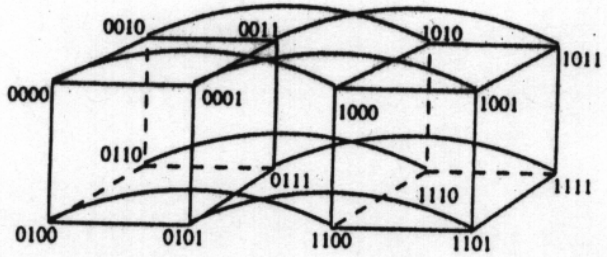
③環立方体網(CCC,Cube Connected Cycle)

$N=n2^n$、次数は３、直径は3n/2

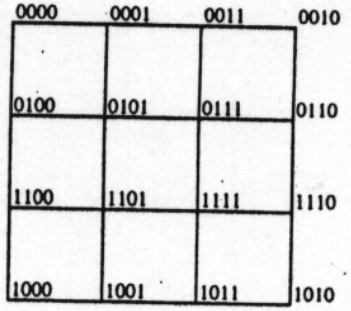
④CCTcube網



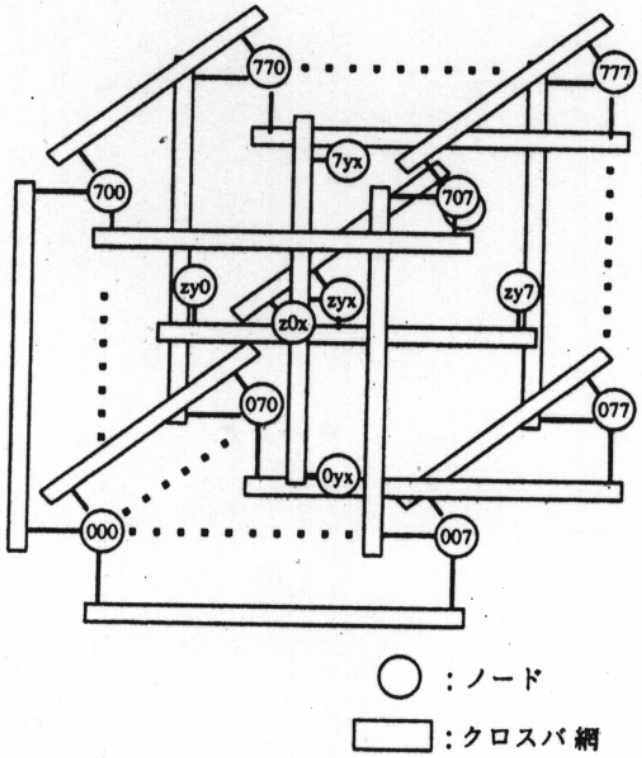
(a)2進 3-キューブ



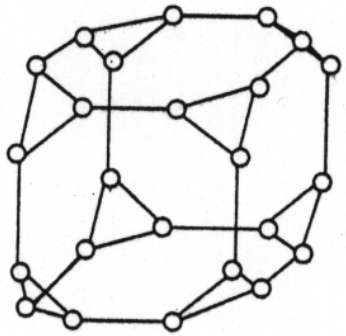
(b) 2進4-キューブの合成



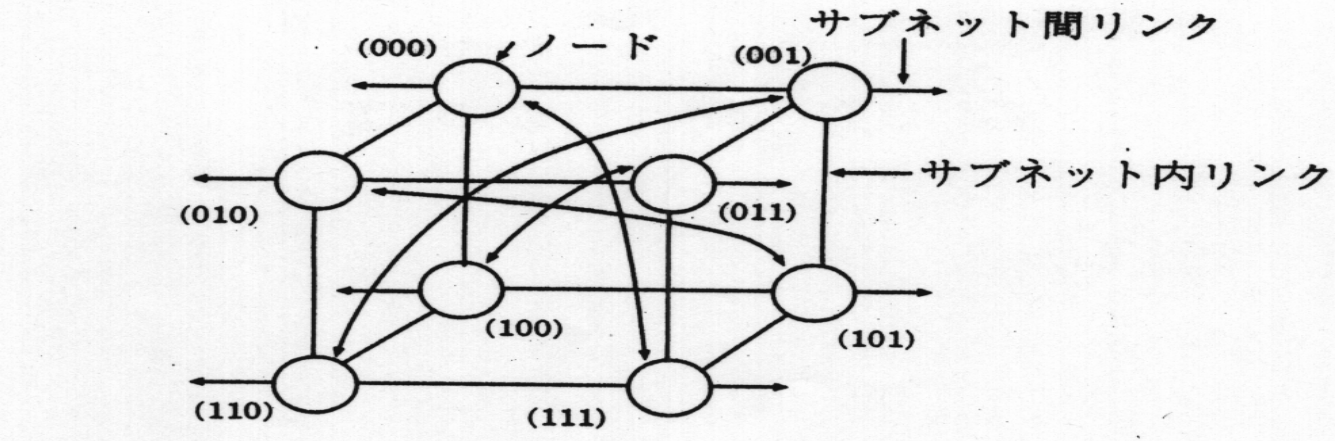
(c) 2進 4-キューブによるトーラスの実現



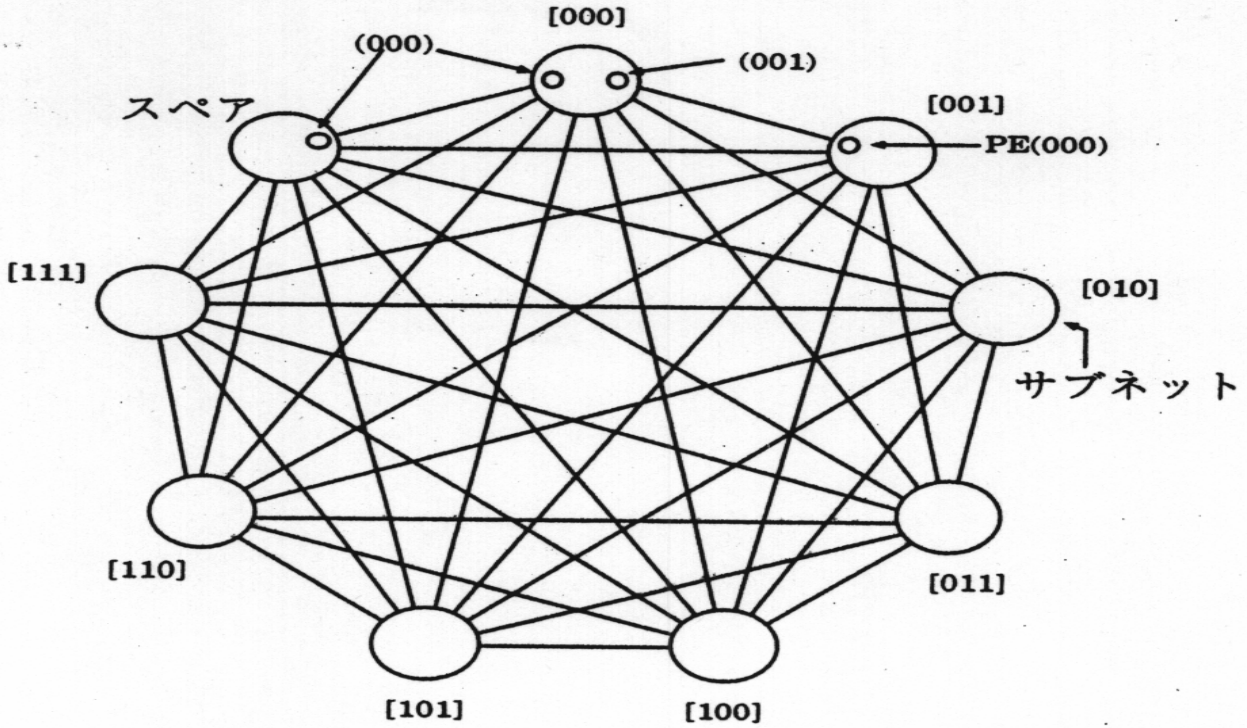
(d) ベース mn-キューブ



(e)環立方体



(a)サブネット



(b)サブネット間の結合

図 2.8　CCT cube の構成

⑤ハイパクロス網

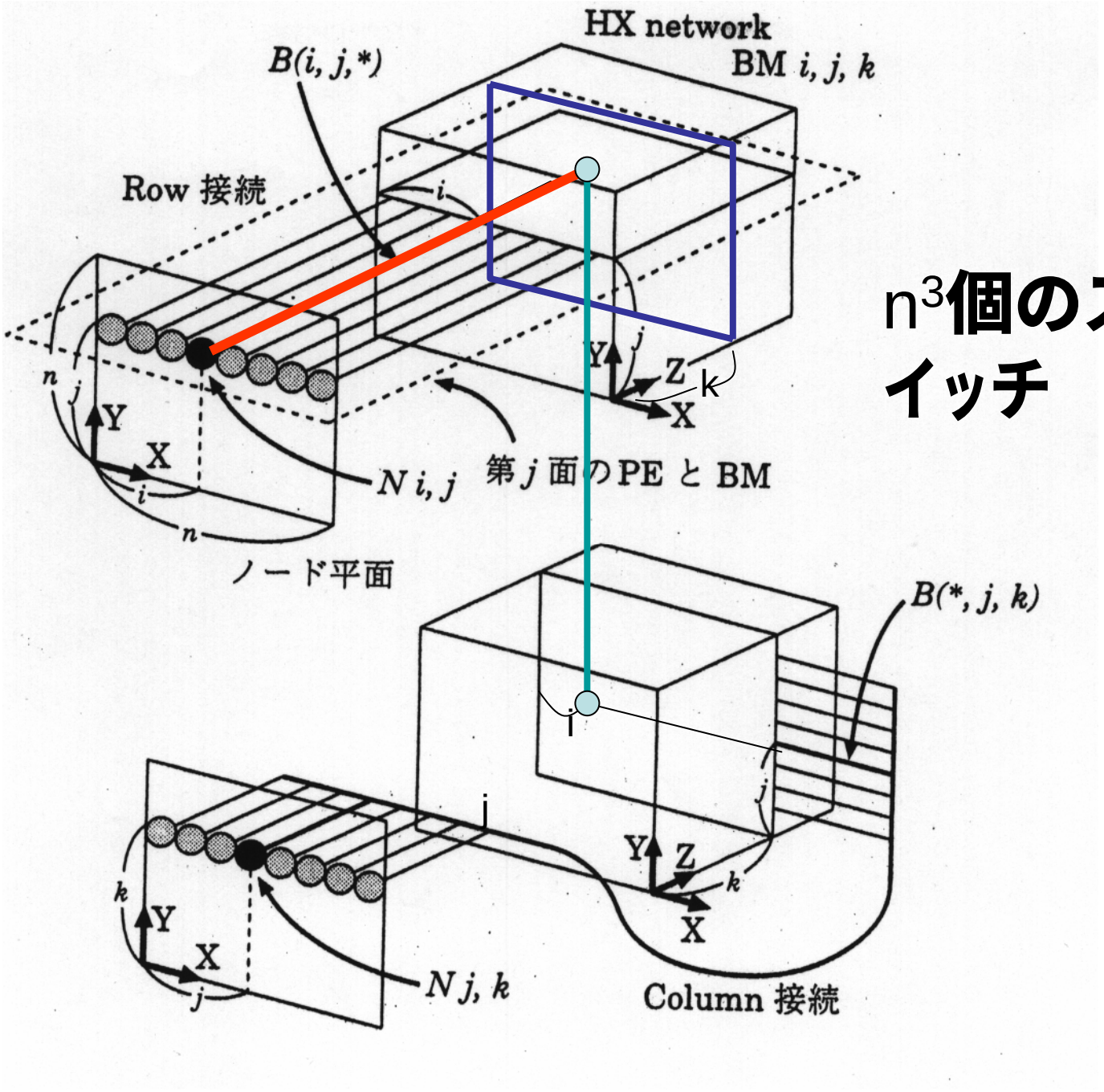
$N=n^2$

ノード番号を$N_{i,j}$

$N_{i,j}$と$N_{j,k}$および$N_{j,k}$と$N_{i,j}$が直接通信

任意のノード間通信$N_{i,j}$と$N_{k,l}$の通信

$N_{i,j}$→$N_{j,k}$→$N_{k,l}$、直径は2


HX network
BM i, j, k
B(i, j, *)
Row 接続
i
n
j
Y
X
i
n
N i, j
第 j 面の PE と BM
Y
Z
X
j
k
ノード平面
B(*, j, k)
i
j
Y
Z
X
k
k
Y
X
j
N j, k
Column 接続

n^3個のスイッチ

・リング網の埋込み

$N_{k,j}(j=1,2,,k)$、$N_{i,k}(i=1,2,,k)$の

$2k-1$個のノード

$(1,k)$、(k,k)、$(k-1,k)$、$(k,k-1)$、

$(k-2,k)$、$(k,k-2)$、$(k-3,k)$、・・・、$(k,1)$

とつなぎ、

$k+1$について

$(1,k+1)$、$(k+1,k+1)$、$(k,k+1)$、$(k+1,k)$、

$(k-1,k+1)$、$(k+1,k-1)$、$(k-2,k+1)$、・・・、

$(k+1,1)$

とつなぐ。

つなぎ目の（$k,1$）と（$1,k+1$）は結合可

・２次元トーラスの埋込み

２次元トラスのノード番号：$T_{x,y}$

$T_{x,y}$---$N_{x,y}$：x+yが偶数のとき

$T_{x,y}$---$N_{y,x}$：x+yが奇数のとき

Ｔ平面でx+yが奇数：

隣接する４つのノードでの

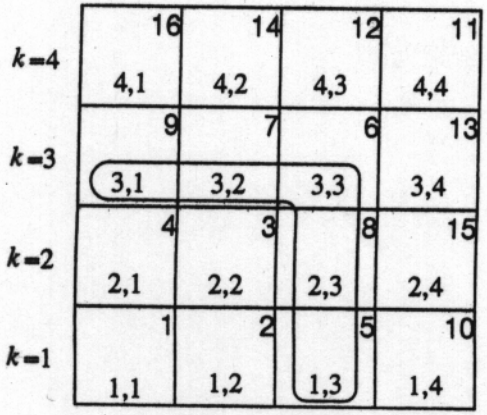
値（X+Y値）はすべて偶数

$T_{x,y}$のみが$N_{y,x}$に写像。

Ｎ平面ではすべてHX結合条件満足

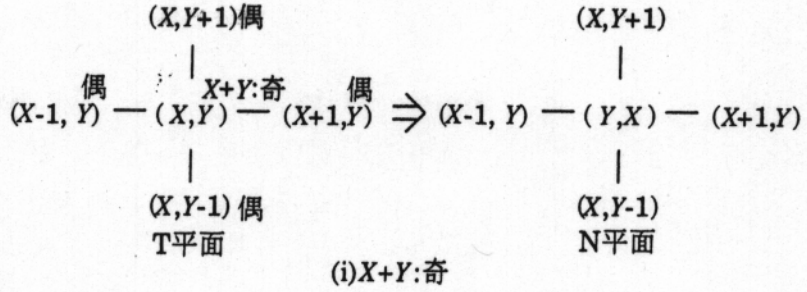
Ｔ平面でx+yが偶数：

$T_{x,y}$以外すべて奇数。

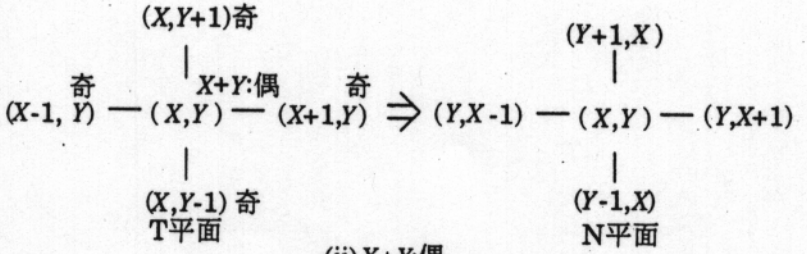


右上角の番号:リング網のノード番号

(a) HX網へのリングの埋込み



(i) $X+Y$:奇



(ii) $X+Y$:偶

4,1	4,2	4,3	4,4
3,1	3,2	3,3	3,4
2,1	2,2	2,3	2,4
1,1	1,2	1,3	1,4

T平面

1,4	4,2	3,4	4,4
3,1	2,3	3,3	4,3
1,2	2,2	3,2	2,4
1,1	2,1	1,3	4,1

N平面

(iii) T平面のN平面への写像

(b) HX網への2次元トーラスの埋込み

・ハイパーキューブの埋込み

$N=2^{2k}$ ($n=2^{k}$)。

kビットの反射２進符号

ある数iとハミング距離１の数j

はk個存在

j-i：奇数

反射2進符号
符号間のハミング距離＝1

	通常2進	反射2進	
	000	000	000 0
1	001	001	001 1
2	010	011	011 2
1	011	010	010 3
3	100	110	110 4
1	101	111	111 5
2	110	101	101 6
1	111	100	100 7

2kビットの数

a(i)、b(i)

2kビット長の符号：a(i)+b(j)

ノード$H_{a(i)+b(j)}$を

ハイパクロス網ノード$N_{x,y}$に写像

x=i、y=j：i+j奇数時

x=j、y=i：i+j偶数時

i+jが奇数のとき

（iが偶数／奇数、jが奇数／偶数）

$H_{a(i)+b(j)}$は$N_{i,j}$に写像

$H_{a(i)+b(j)}$と結合しているのは、

b(j)を固定して考えると

k個の$H_{a(r)+b(j)}$、r-i：奇数

a(i)を固定して考えると

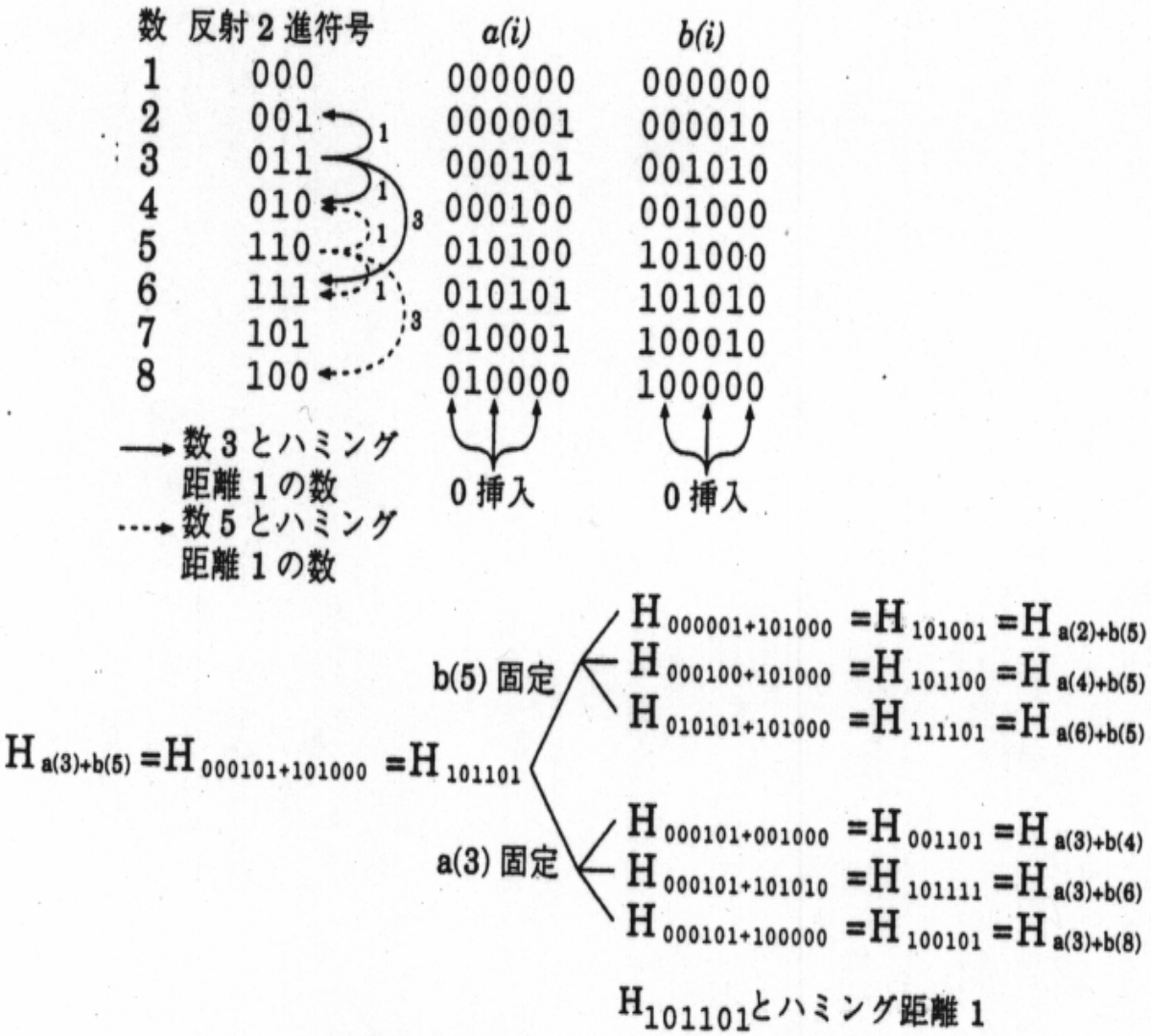
k個の$H_{a(i)+b(s)}$ 、s-j：奇数

（i:偶、j:奇、r:奇、s:偶）または

（i:奇、j:偶、r:偶、s:奇））

であり，r+j、i+sは偶数

ノードは$N_{j,r}$、$N_{s,i}$に写像：$N_{i,j}$と結合



(c) HX 網へのハイパキューブの埋込み

図 2.10　HX 網への各種網の埋込み

（7）シャフル網系

①シャフル・エクスチェンジ網

②de Bruijn 網

次数：４、直径：$\mathrm{Log}_2 N$

ハミルトンパス：

000,001,011,111,110,101,010,100

リング網の埋込み

nビット列： $X_i=(x_{i1},x_{i2},\ldots,x_{in})$

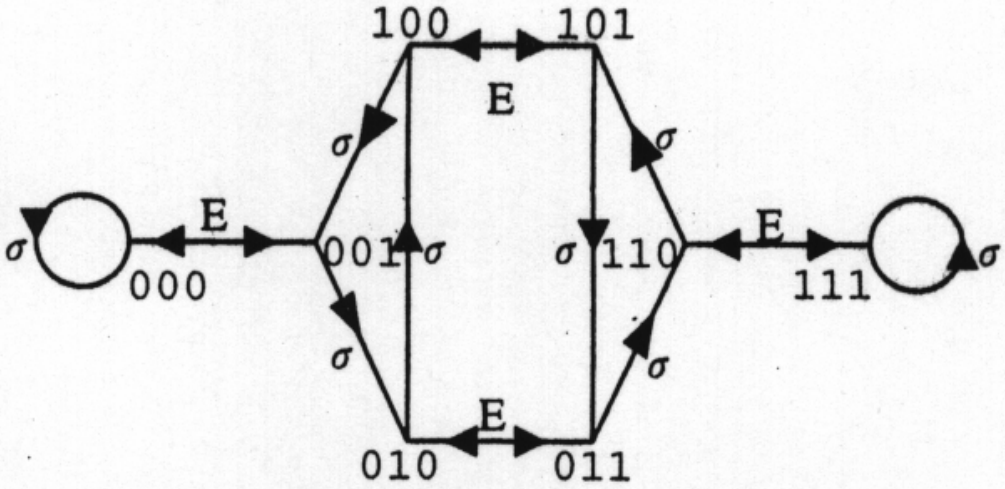
X_i**の共役nビット列：** $X_i{}'=(\sim x_{i1},x_{i2},\ldots,x_{in})$

（X_1**,** $X_2,\ldots,X_k$**）：サイクル⇔**

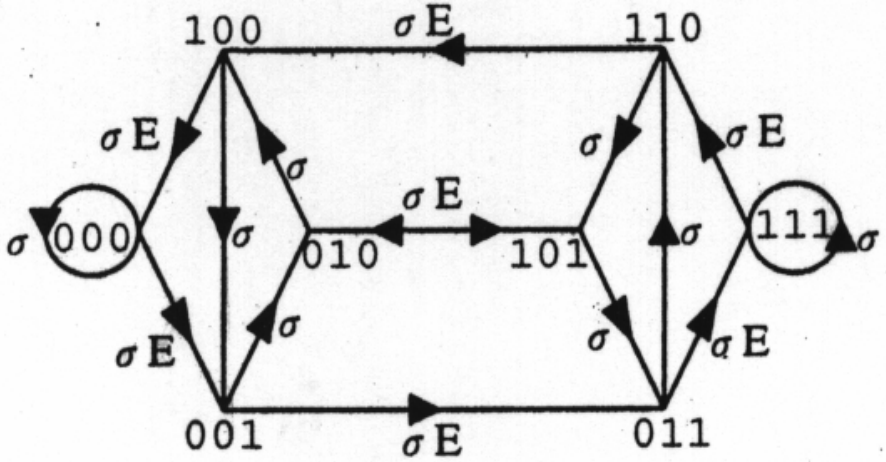
$\sigma X_1=X_2$**,** $\sigma X_2=X_3,\ldots,$ $\sigma X_{k-1}=X_k$**,** $\sigma X_k=X_1$

$\sigma X_i=\sigma EXi'$

$\sigma EX_i=\sigma X_i'$



(a) シャフル・エクスチェンジ網



(b) de Bruijn網

相互隣接サイクル　C_1，C_2

$C_1 = (X_1^1, X_2^1, \ldots, X_k^1)$，$C_2 = (X_1^{1'}, X_2^2, \ldots, X_m^2)$

C_1, C_2を１つの拡張サイクルに合成可能

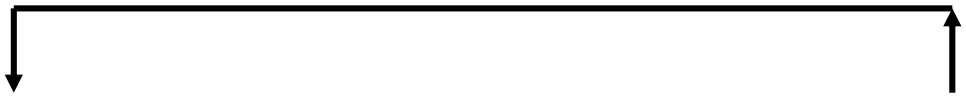
共役

拡張サイクル： σE**使用可能**

$\sigma X_1^1 = \sigma E X_1^{1'} = X_2^1$

$\sigma E X_1^1 = \sigma X_1^{1'} = X_2^2$

C_1**サイクル**



X_1^1　　$X_2^1 \rightarrow \sigma \cdots \rightarrow \sigma \cdots X_k^1$

$\downarrow \sigma E$　　$\uparrow$　　σE

$X_2^2 \rightarrow \sigma \rightarrow X_3^2 \rightarrow \sigma \cdots \rightarrow \sigma \cdots X_m^2 \rightarrow \sigma \rightarrow X_1^{1'}$

C_2**サイクル**

重み　$W(X_i)=\sum_{j=1}^{n} x_{ij}$

アルゴリズム

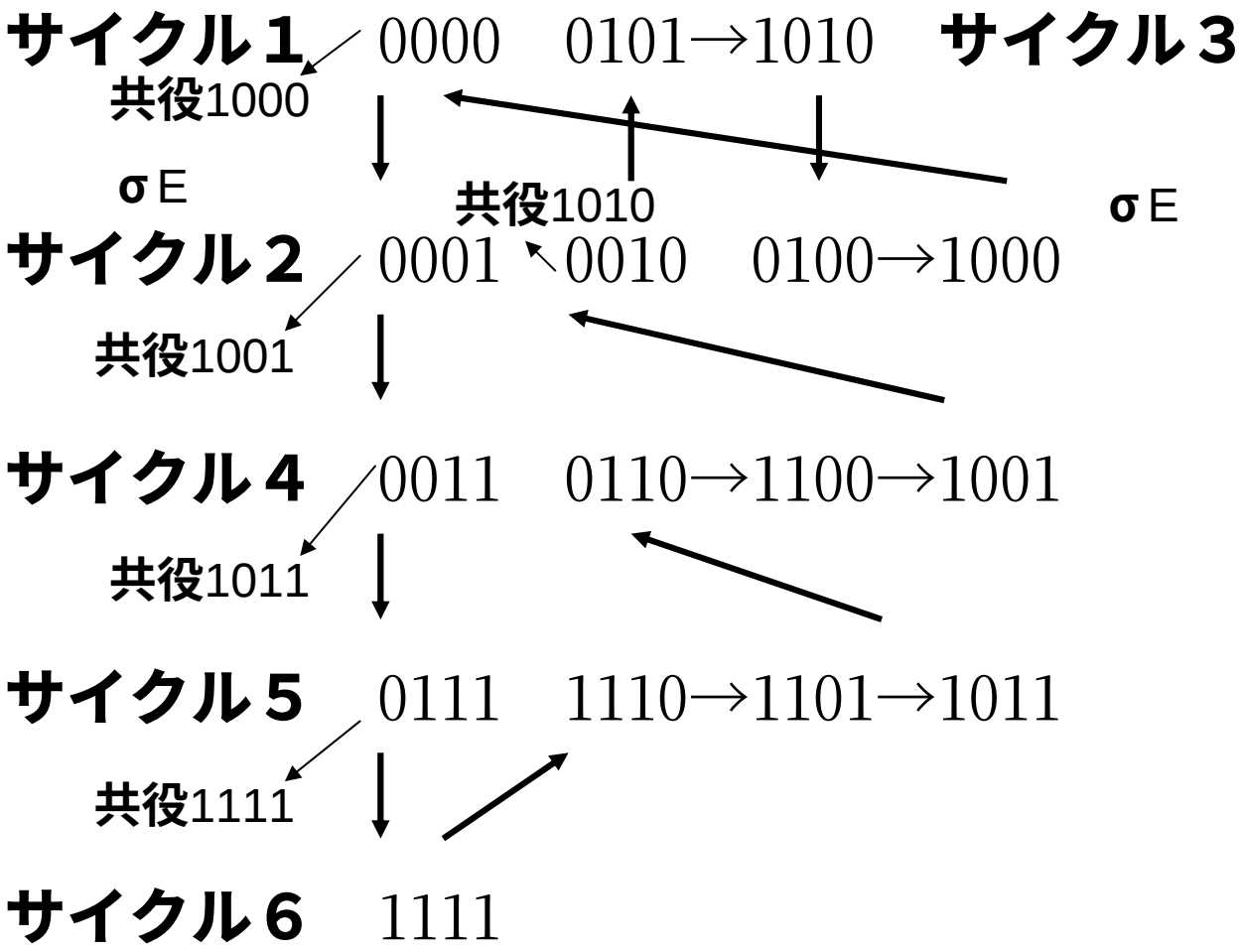
(1)**重み０のｎビット列**(0,0,...,0)**からスタート**

(2)**重み**K-1**以下のサイクルをマージした拡張サイクル**EC^{K-1}**の要素**X**と共役な**X'**を要素に持つ重み**K**のサイクルを選びマージし，**EC^{K}**とする．**

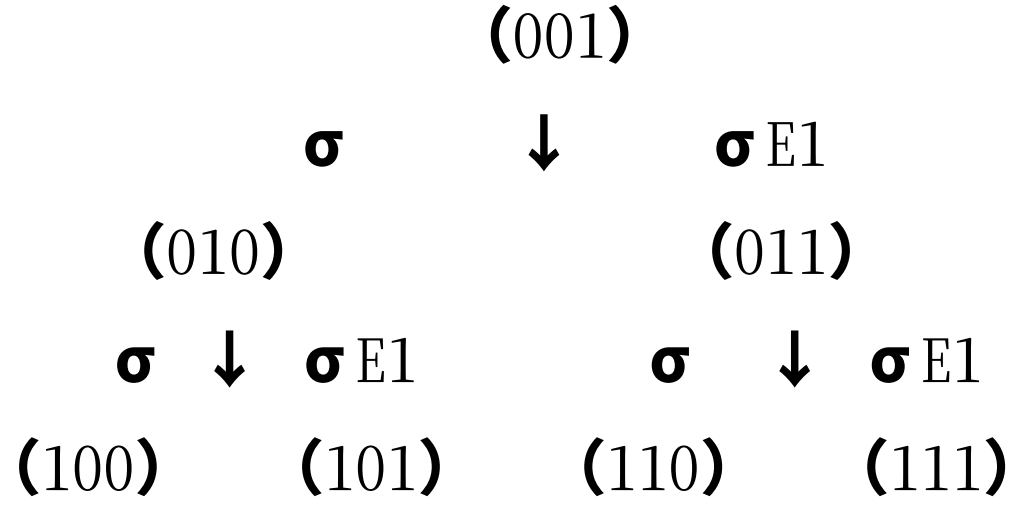
(3)K=n**まで**(2)**を繰り返す．**

例

重み０	**サイクル１**	0000
重み１	**サイクル２**	0001→0010→0100→1000
重み２	**サイクル３**	1010→0101
	サイクル４	0011→0110→1100→1001
重み３	**サイクル５**	0111→1110→1101→1011
重み４	**サイクル６**	1111


サイクル１
0000
0101→1010
サイクル３
共役1000
σE
共役1010
σE
サイクル２
0001
0010
0100→1000
共役1001
サイクル４
0011
0110→1100→1001
共役1011
サイクル５
0111
1110→1101→1011
共役1111
サイクル６
1111

トリー網の埋込み



トーラス網：不可

de Bruijn**網の一般形：**

D進n桁$(D_n, D_{n-1}, \ldots, D_1)$**が**

↓$(D_{n-1}, D_{n-2}, \ldots, D_1, *)$**と結合**

③循環オメガ網（circular omega）

オメガ網のスイッチをノードで置き換えた網（後述）

④Star　Graph網

ノード数N:n!

記号列の最初の文字と他の文字を

入れ換えたノードとリンク

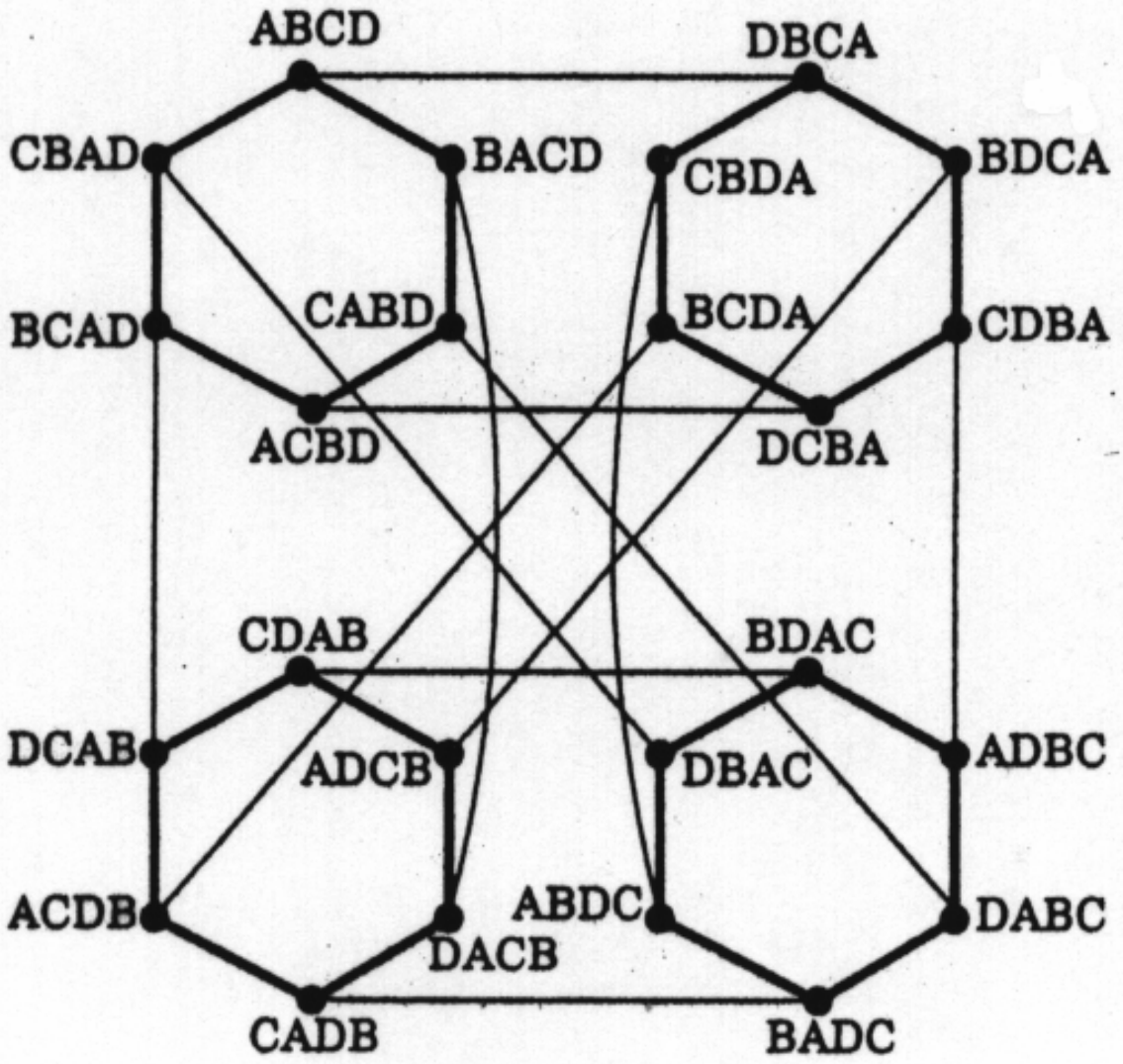
ノードABCD：BACD、CBAD、DBCAと結合

次数：n-1

任意のノード番号から

ノードABCDEFGHへのルーティング

記号A,B,C,D,E,F,G,Hの



(c)スターグラフ網の構成

ホームポジション：1,2,3,4,5,6,7,8

（たとえば、Aは1、Cは3、Hは8**）**

- **与えられた送信ノードの最初の記号が**Aでない

  その記号に対応する送信ノード番号中のホームポジションにある記号と置き換えを行う。

- **与えられた送信ノードの最初の記号が**A

  適当な位置にある記号と入れ換える。

- **以上の過程を繰り返す。**

FDGBEHCA　HDGBEFCA　ADGBEFCH　DAGBEFCH

BAGDEFCH　ABGDEFCH　GBABEFCH　CBADEFGH

→ABCDEFGH

記号列のサイクル：サイクル内の入れ換えホームポジションへ

FDGBEHCA：

(F,H,A)と(D,B)と(G,C)サイクルの形成

ルーティング回数N_r

N_r=C+m-2：送信ノード番号の先頭がAでないとき

C+m　：送信ノード番号の先頭がAのとき

FDGBEHCAの例：C=3,m=7、N_r:8

①サイクルの決定

FDGBEHCAでは、サイクル１：（F,H,A）、

サイクル1：(F,H,A)

サイクル２：（D,B）、サイクル３として：（G,C)

②サイクル１での置換

FDGBEHCA－＞

HDGBEFCA－＞ADGBEFCH

サイクル１：サイクル長３、

２回の置換でADGBEFCH：一般にもとのサイクル長-1

③サイクル２に統合

ADGBEFCH－＞DAGBEFCH－＞

BAGDEFCH－＞ABGDEFCH

Ａをサイクル２に統合：サイクル長３、

３回の置換でABGDEFCH：一般にもとのサイクル長＋１

④サイクル３に統合：

ABGDEFCH－＞GBADEFCH－＞

CBADEFGH－＞ABCDEFGH

サイクル３、サイクル長３、

３回の置換でABCDEFGH：一般にもとのサイクル長＋１

先頭がAでないとき

置換回数：m-1+C-1でC+m-2

先頭がAのとき：C+m

C-1個のサイクルで＋1の置換

最初のサイクルで置換-1

直径

先頭：A

サイクル：（D,B,C）、（F,E）、（H,G）：$\lfloor (n-1)/2 \rfloor$

サイクル長：総数がn-1のとき

$$\lfloor 3(n-1)/2 \rfloor$$

２．３静的網のルーティング、デッドロック

２.３.１パケット交換方式

①蓄積交換（store and forward）

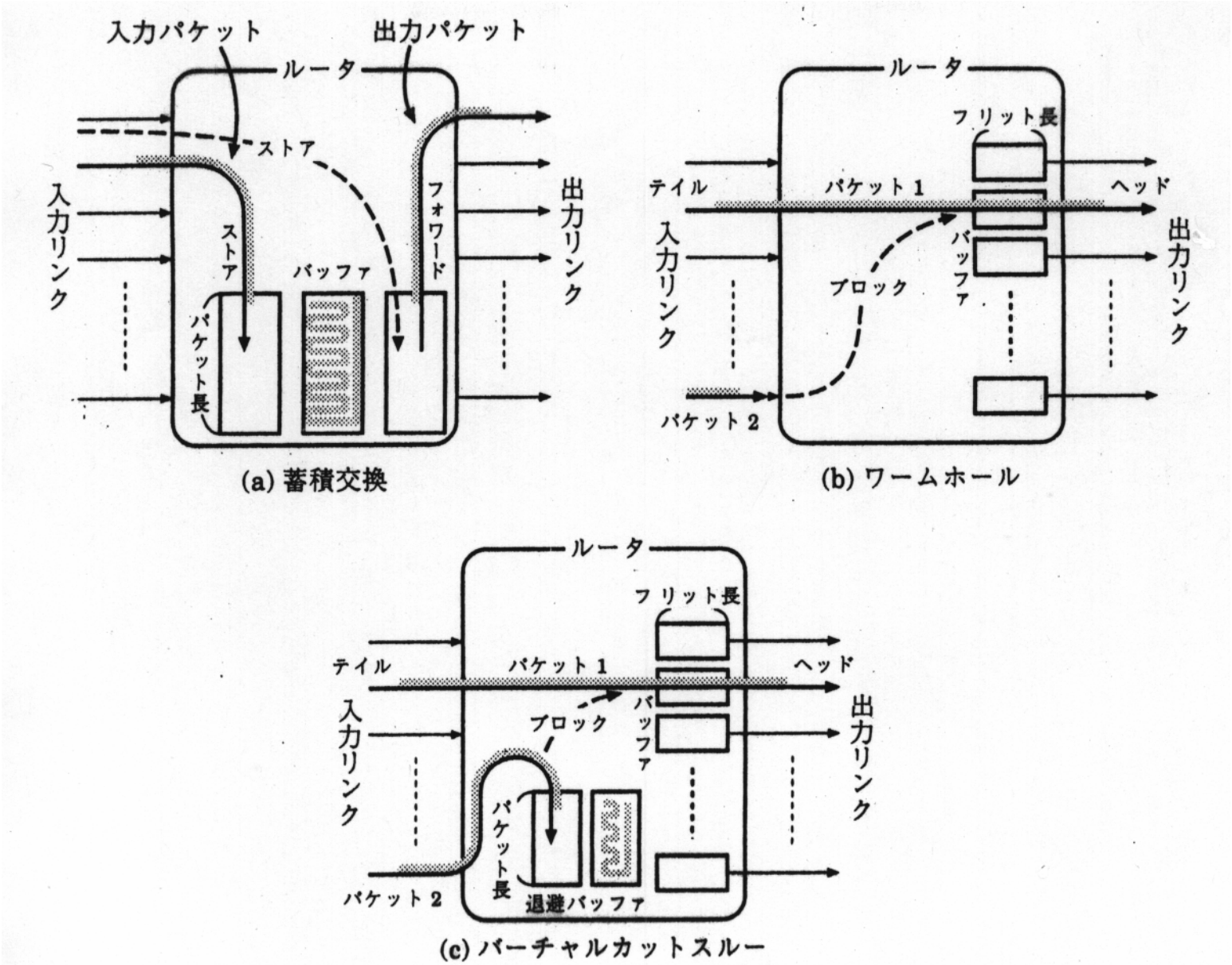
②ワームホール

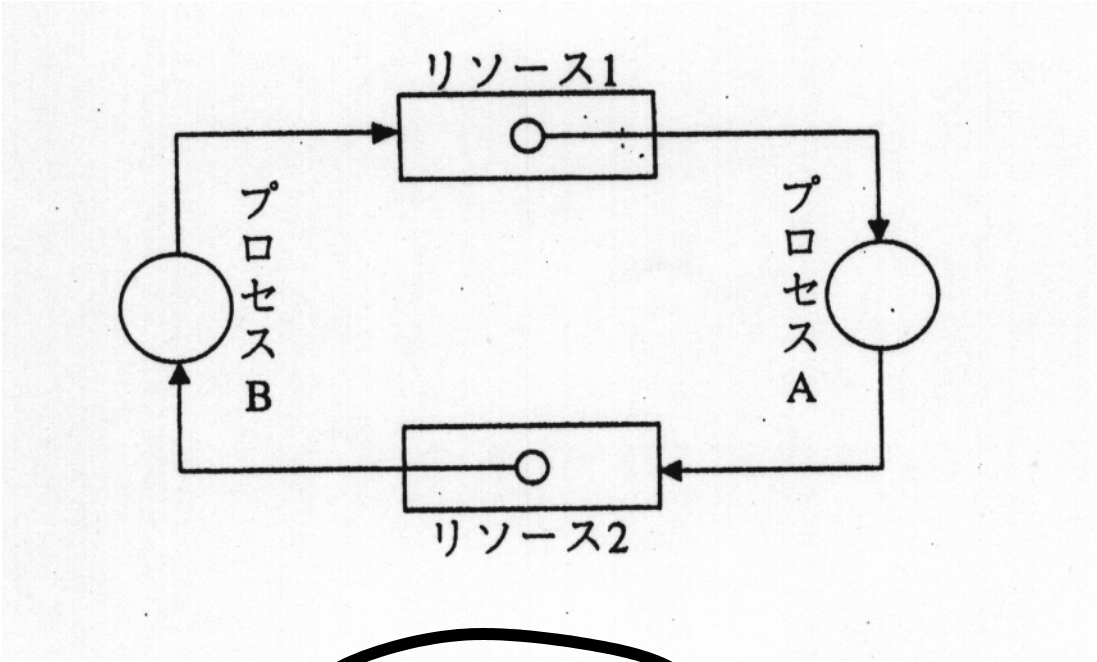
③バーチャルカットスルー

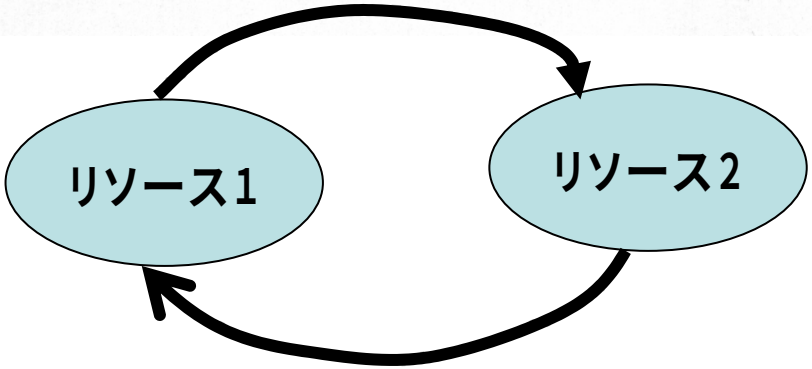
２.３.２デッドロック一般論

デッドロックの発生条件

①排他的利用（Exclusive control）


入力パケット
出力パケット
ルータ
ストア
入力リンク
出力リンク
ストア
バッファ
フォワード
パケット長
(a) 蓄積交換
ルータ
フリット長
テイル
パケット1
ヘッド
入力リンク
出力リンク
バッファ
ブロック
パケット2
(b) ワームホール
ルータ
フリット長
テイル
パケット1
ヘッド
入力リンク
出力リンク
バッファ
ブロック
パケット長
パケット2
退避バッファ
(c) バーチャルカットスルー


リソース1
プロセスB
プロセスA
リソース2


リソース1
リソース2

②非妥協的リソース待ち（hold-wait）

③横取り不可（non-preemption）

④リソース利用におけるサイクルの存在

デッドロック対処手段

（１） デッドロック検出（detection）

（２） デッドロック回避（avoidance）

銀行家のアルゴリズム

将来デッドロックが生じる可能性のある

リソース割付けの禁止

システム状態：安全（safe）状態に保つ

安全な状態にあるプロセス：

すべて実行終了可能

$P_0, P_1, P_2, \ldots, P_m$**プロセス：**

リソース利用最大値(M_i)**が既知**

システムが安全状態

現在リソース利用値：U_i

現在リソース要求残値$B_i : (M_i - U_i)$

現在リソース要求値：R_i

空きリソース数T**：**

（全リソース数-ΣU_i**）**

$T >= R_i$

	A社	銀行	B社
	限度額 20 万円	資本 20 万円	限度額 15 万円
1 月	5 万円要求		
	受理	15 万円	
2 月			8 万円要求
		7 万円	受理
3 月	4 万円要求		
	受理	3 万円	
4 月			1 万円要求
		2 万円	受理
5 月	3 万円要求		
	拒否		
6 月		3 万円要求	
		拒否	

A, B 両社とも事業の推進不可能

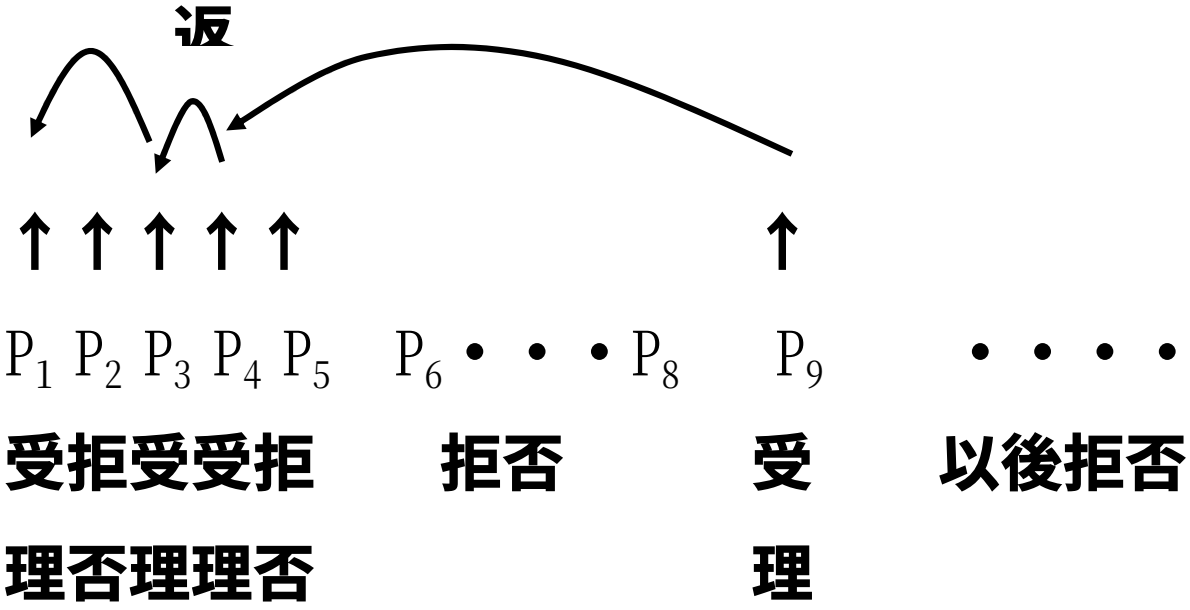
$T>=B_i$

一方，$T \geqq B_i$ のとき受理される方式では次のようにデッドロックは生じない。

	A社	銀行	B社
	限度額 20 万円	資本 20 万円	限度額 15 万円
1 月	5 万円要求		
	受理	15 万円	
2 月			8 万円要求
		7 万円	受理
3 月	4 万円要求		
	拒否（将来の予想借入れ最大額 15 万円＞銀行資金残高 7 万円）		
4 月			1 万円要求
		6 万円	受理
5 月			3 万円要求
		3 万円	受理
6 月			3 万円要求
		0 万円	受理
7 月		15 万円	15 万円返済

P_iのリソース要求

$T \geqq B_i$のとき要求を受理



（３）デッドロック防止（prevention）

①リソースの共有化

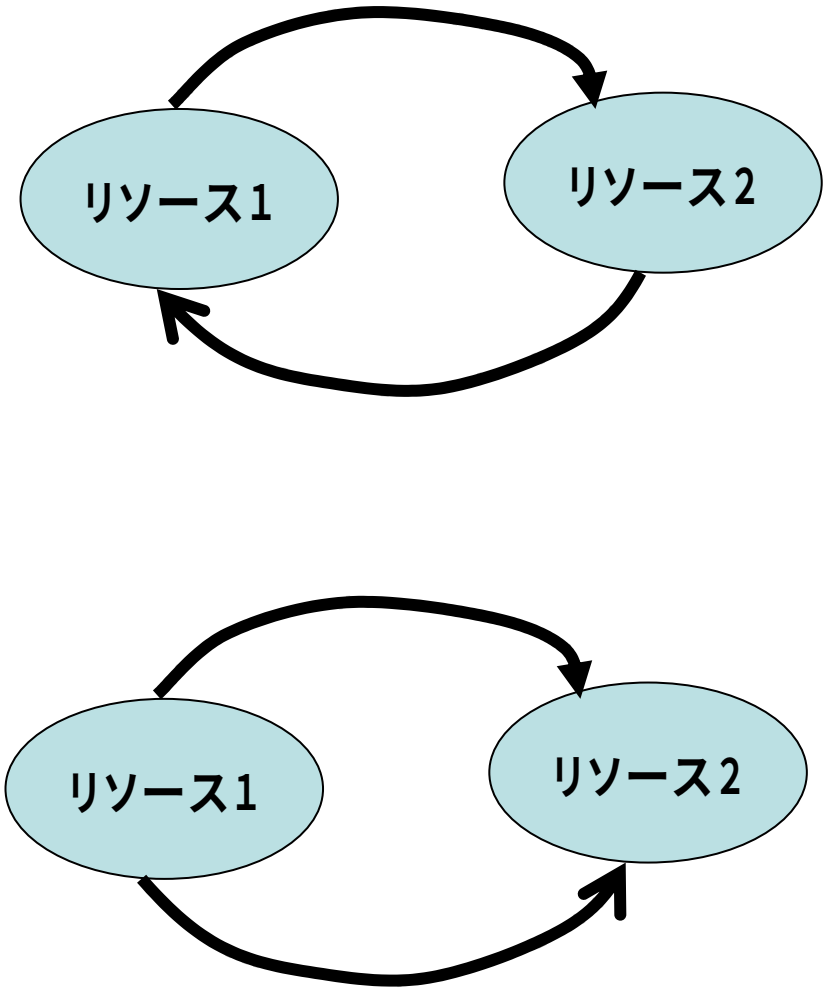
②妥協的リソース待ち

- 非受理時全リソース解放
- 一括リソース要求

③横取り可能

④サイクルの防止

- リソース数の増大
- 代替リソースの提供
- リソース番号順獲得


リソース1
リソース2
リソース1
リソース2

２．３．３デッドロック

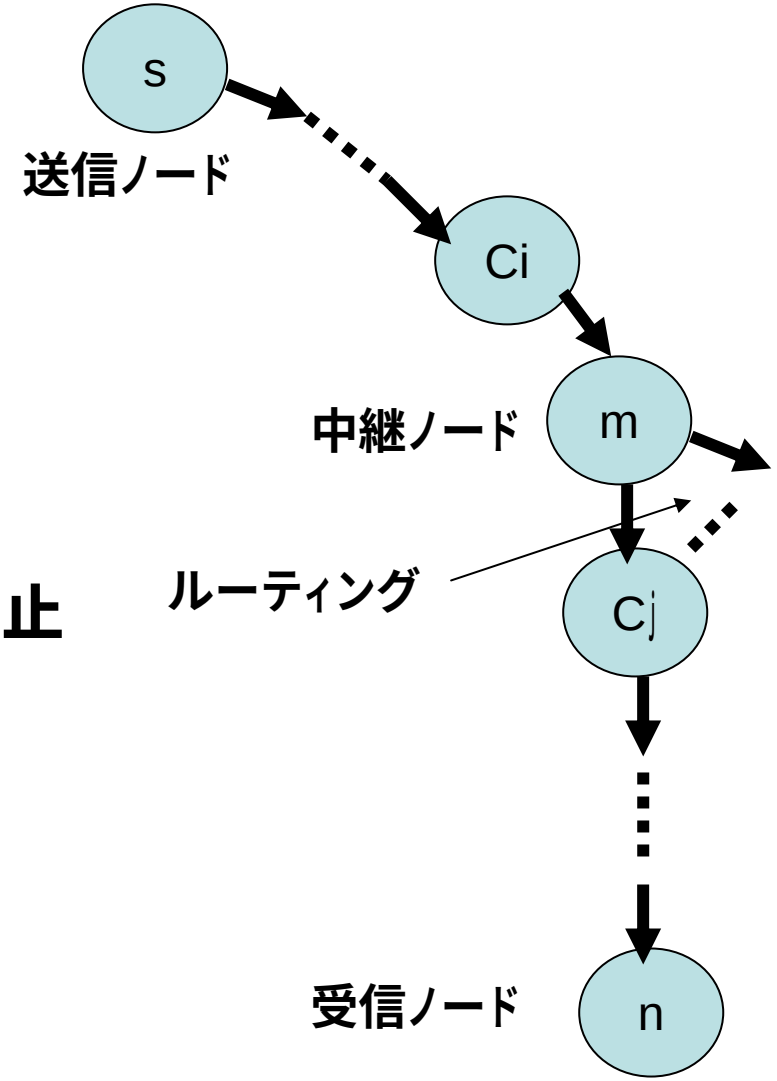
（１）チャネル依存グラフ

チャネル依存グラフ$D=G(C,E)$

$E=\{(C_i,C_j)\mid R(C_i,n)=C_j\}$

R：ルーティング方式

n：宛先ｎ



（２）リング網でのデッドロック防止

①チャネルの共有化

仮想化につながる考え

②妥協的チャネル待ち

バーチャルカットスルーの考え方

退避バッファ：デッドロック問題

③横取り可能チャネル

④サイクルの防止

・リングの切断

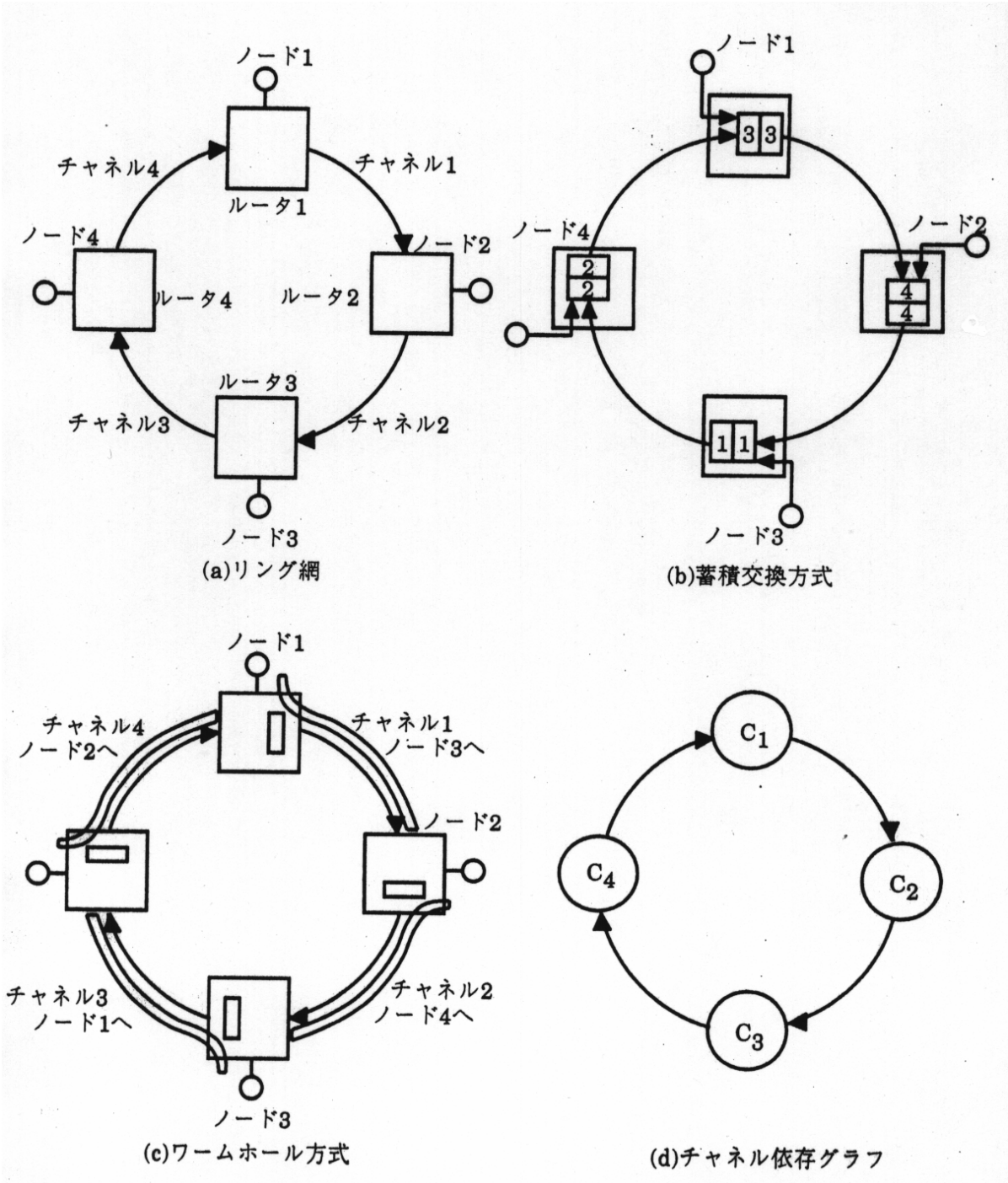
・仮想チャネル

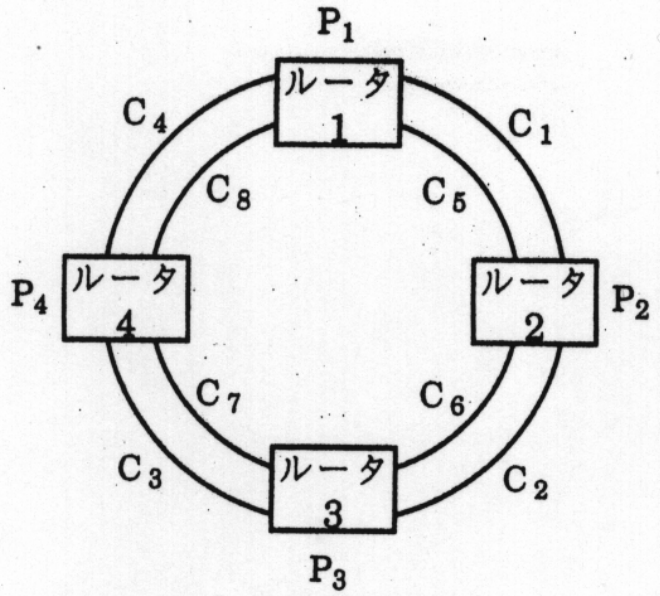
仮想チャネル

一意の番号付け

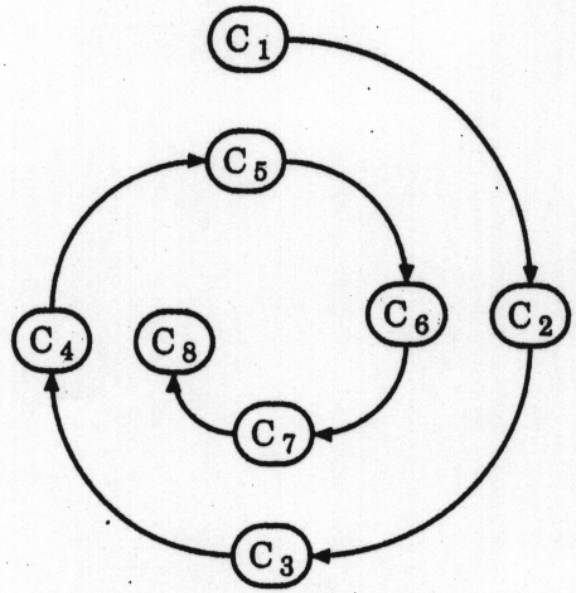
（チャネル番号：C_1、C_2，．．．，C_8）

C_1からC_4：下位リング

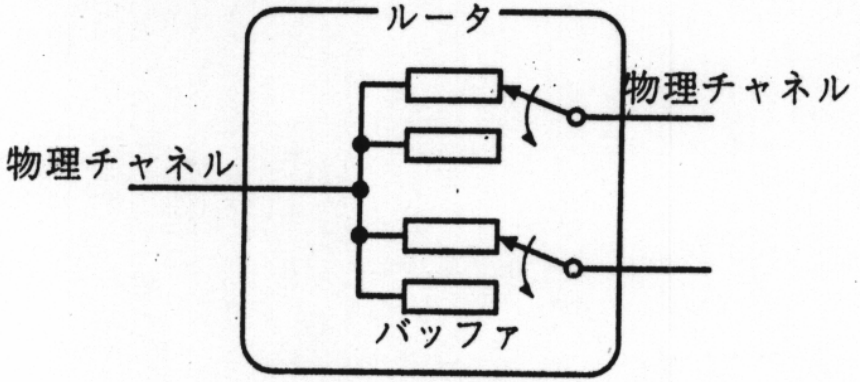

ノード1
チャネル4
チャネル1
ルータ1
ノード4
ノード2
ルータ4
ルータ2
ルータ3
チャネル3
チャネル2
ノード3
(a)リング網
ノード1
3 3
ノード4
ノード2
2
2
4
4
1 1
ノード3
(b)蓄積交換方式
ノード1
チャネル4
ノード2へ
チャネル1
ノード3へ
ノード2
チャネル3
ノード1へ
チャネル2
ノード4へ
ノード3
(c)ワームホール方式
C1
C4
C2
C3
(d)チャネル依存グラフ



(a) 仮想チャネル



(b) チャネル依存グラフ



(c) ルータの構成

C_5からC_8：上位リング

ルーティング方式：チャネル番号が増大する方向

プロセッサ１から３への通信：C_5とC_6

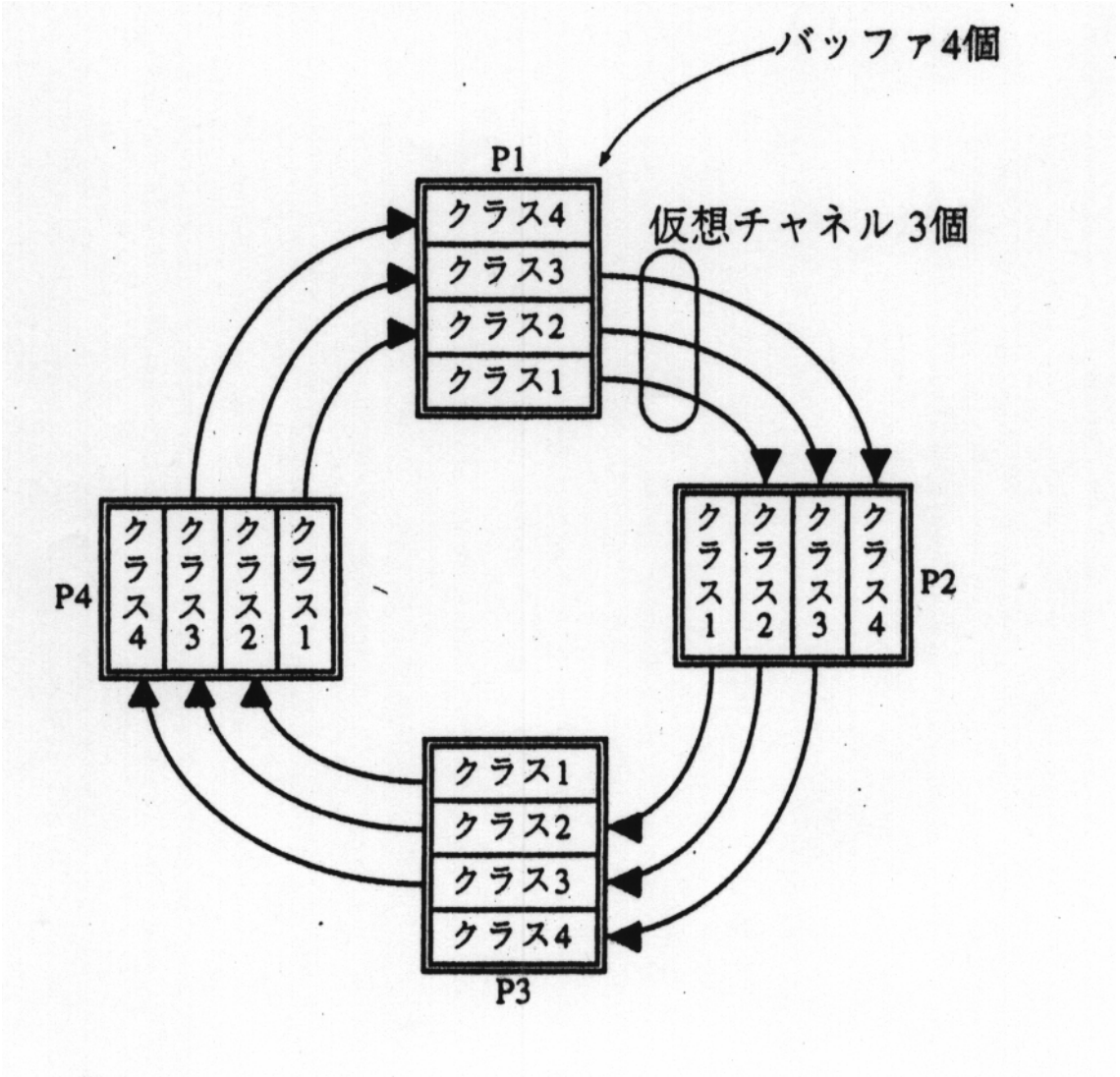
プロセッサ４から２への通信：C_4とC_5

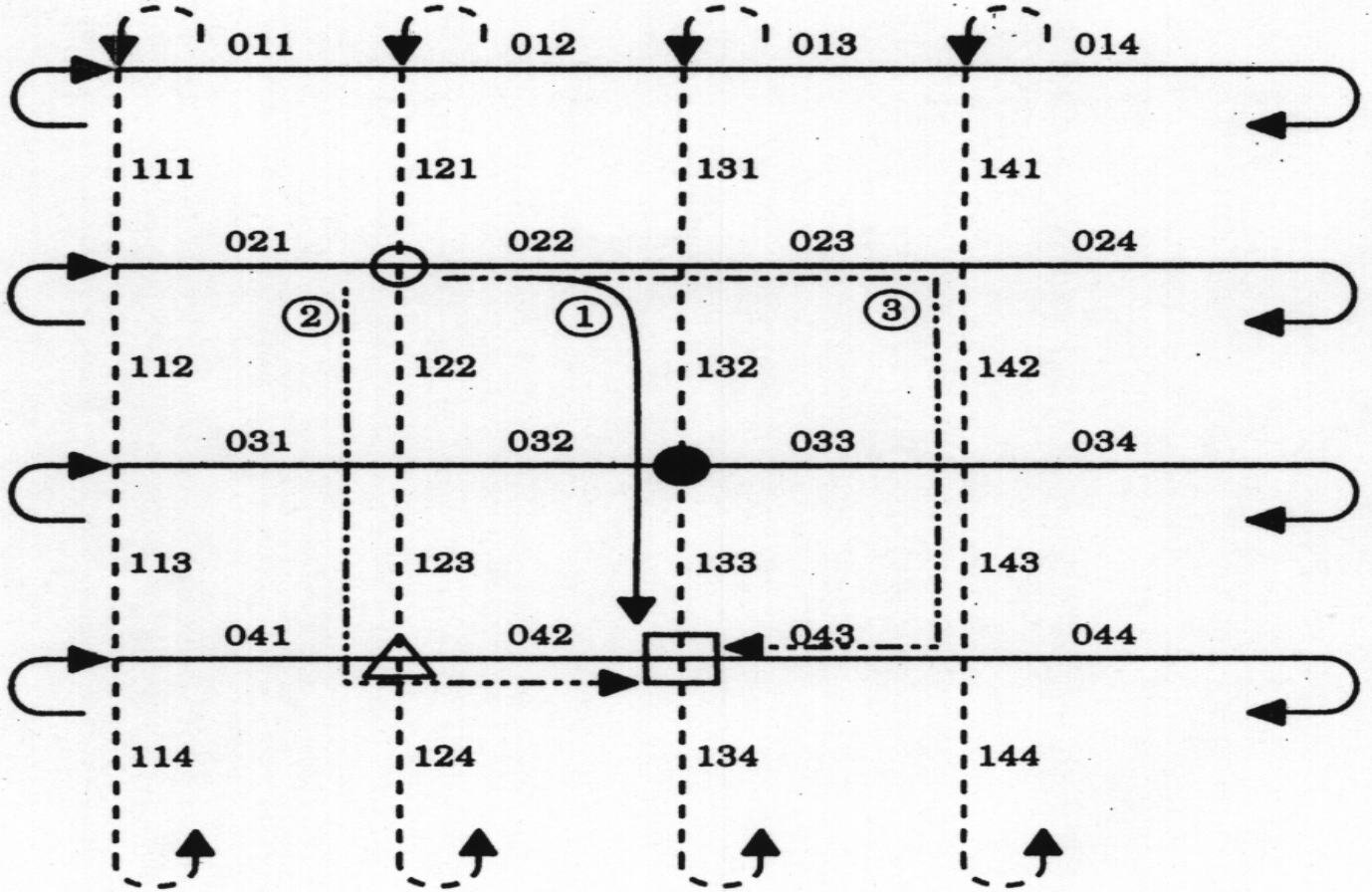
ルータ番号mに到着したパケットの受信先番号mより小さいとき、下位リングを、

大きいときは上位リングを選択

（３）トーラス網、ハイパキューブ網でのデッドロック防止

次元順ルーティング


バッファ4個
P1
クラス4
クラス3
クラス2
クラス1
仮想チャネル 3個
P2
クラス1
クラス2
クラス3
クラス4
P3
クラス1
クラス2
クラス3
クラス4
P4
クラス4
クラス3
クラス2
クラス1



実線　：　X方向
点線　：　Y方向
①　：　次元順ルーティング
②　：　次元逆順ルーティング
③　：　迂回ルーティング

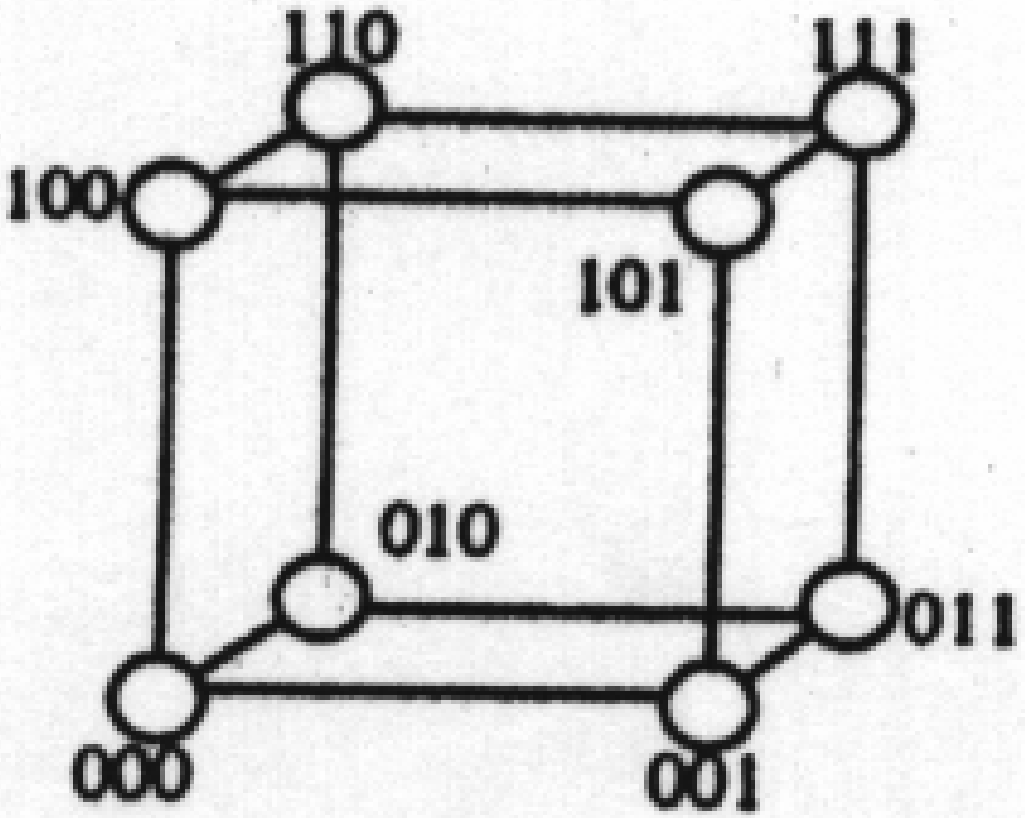
①２次元トーラスの場合

②ハイパーキューブの場合

次元順ルーティングでのデッドロック防止

すべてのチャネルに一意の番号付けができ、

昇順に従ったチャネル番号の獲得



(a)2進 3-キューブ

2.3.4動的ルーティング

次元順ルーティング

デッドロック防止が可能

ルーティングが固定方式

トラフィックの偏り、耐故障に弱い

①デッドロックを生じない方式

②受信ノードに少なくとも辿り着く方式

(1) 仮想チャネルを使用した方法

(a) シンプルな方法

0→15のとき： X+Y-ネット

10→0のとき：X-Y-ネット

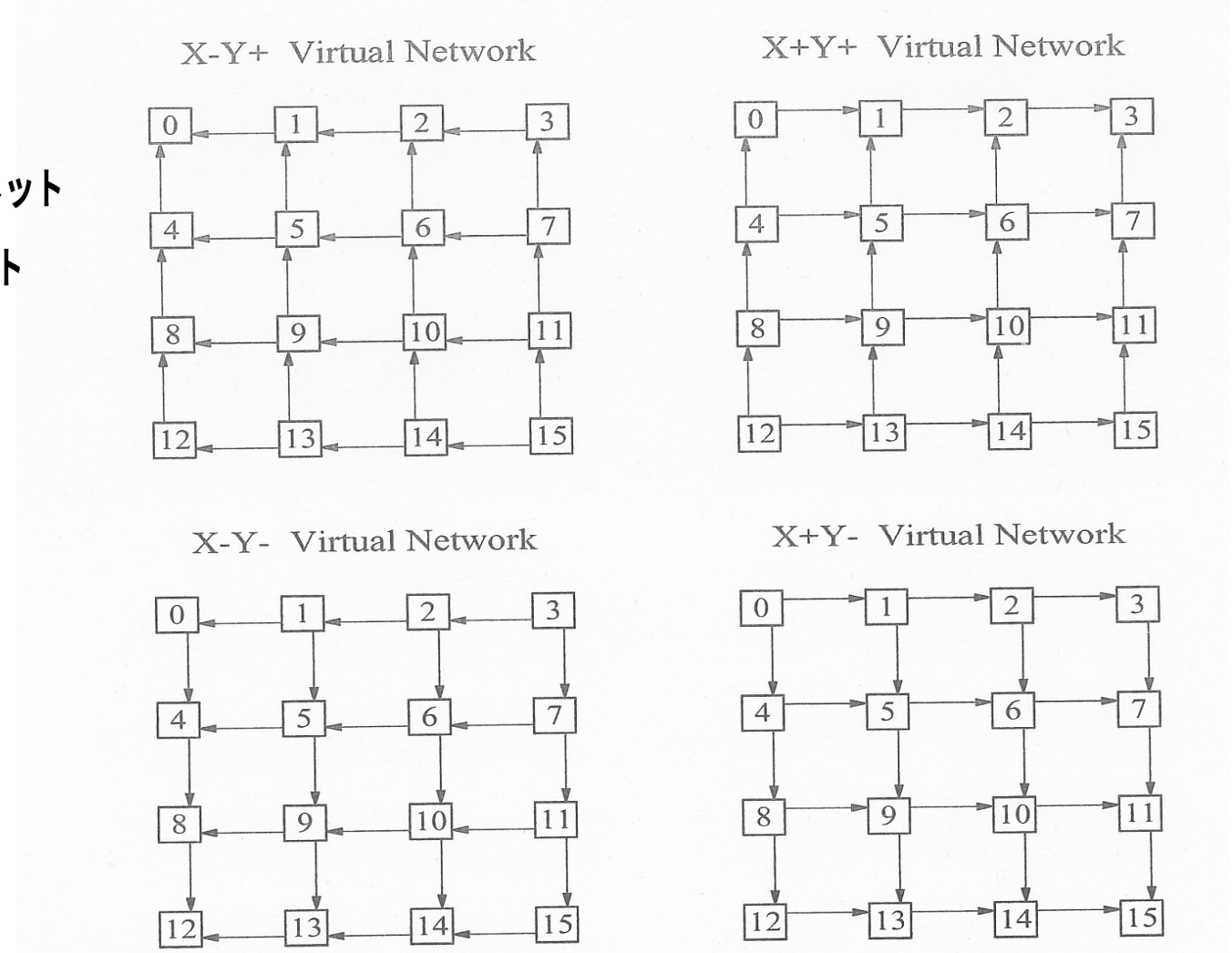


Figure 4.16. Virtual networks for a 2-D mesh.

（ｂ）逆次元順ルーティング

ルータ間のチャネル：

仮想チャネルによってｒ多重

ｒ個のサブネットワークを構成：

クラスｉサブネットワーク

パケットにクラスフィールドC：初期値を０

パケット：クラス番号Cに対応した

サブネットワーク内で任意の次元方向

（Ｘ方向、Ｙ方向）にルーティング

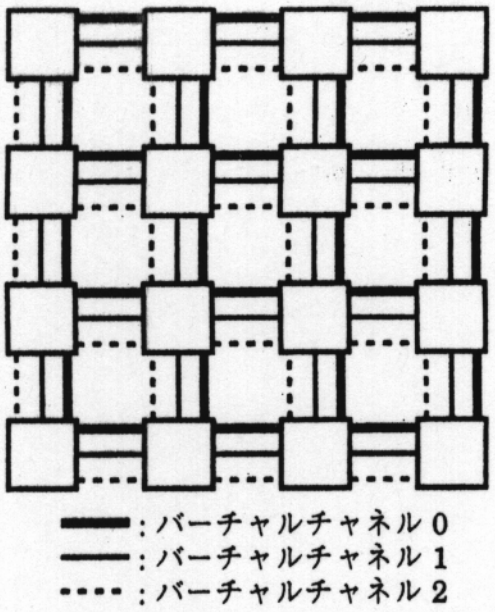
高位次元から低位次元へルーティングの切換え

が生じたとき（Ｙ方向後、Ｘ方向に切り替え）、

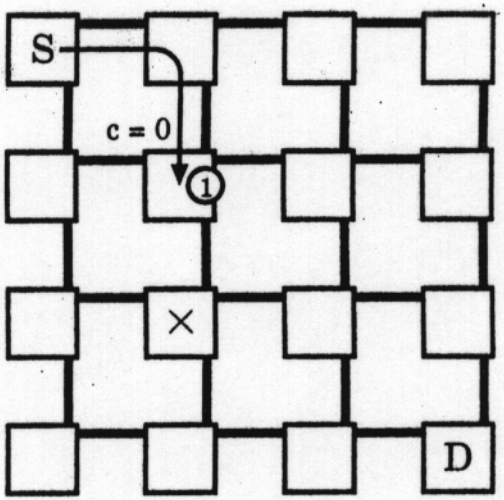
Ｃ値を＋１

Ｃ=ｒのとき以後はクラスｒのサブネット

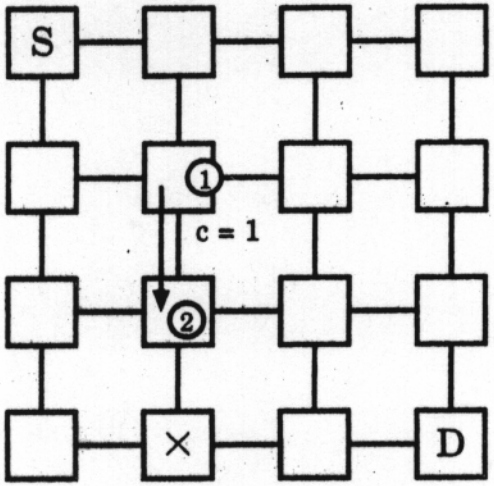
ワークで次元順



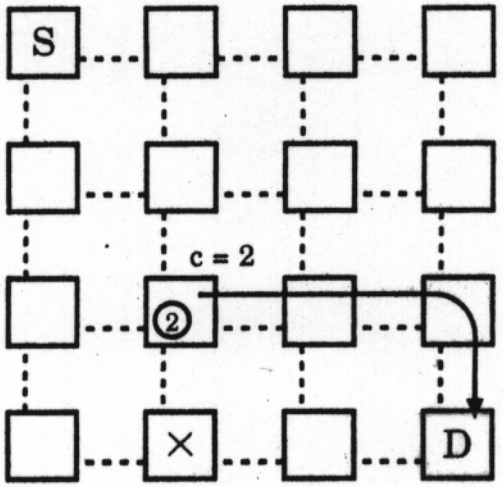
(a) バーチャルチャネル



(b) クラス 0 サブネットワーク



(c) クラス 1 サブネットワーク



(d) クラス 2 サブネットワーク

(2) BlueGene/Lの方式

Escape Ring: **単方向、次元順ルーティング,**

外部からQueueに2つ空きの時:注入可能

内部からQueueに1つの空きの時:投入可能

デッドロックなし

Adaptive Ring:**Queueが一杯の時,** Escape Ring**に注入**

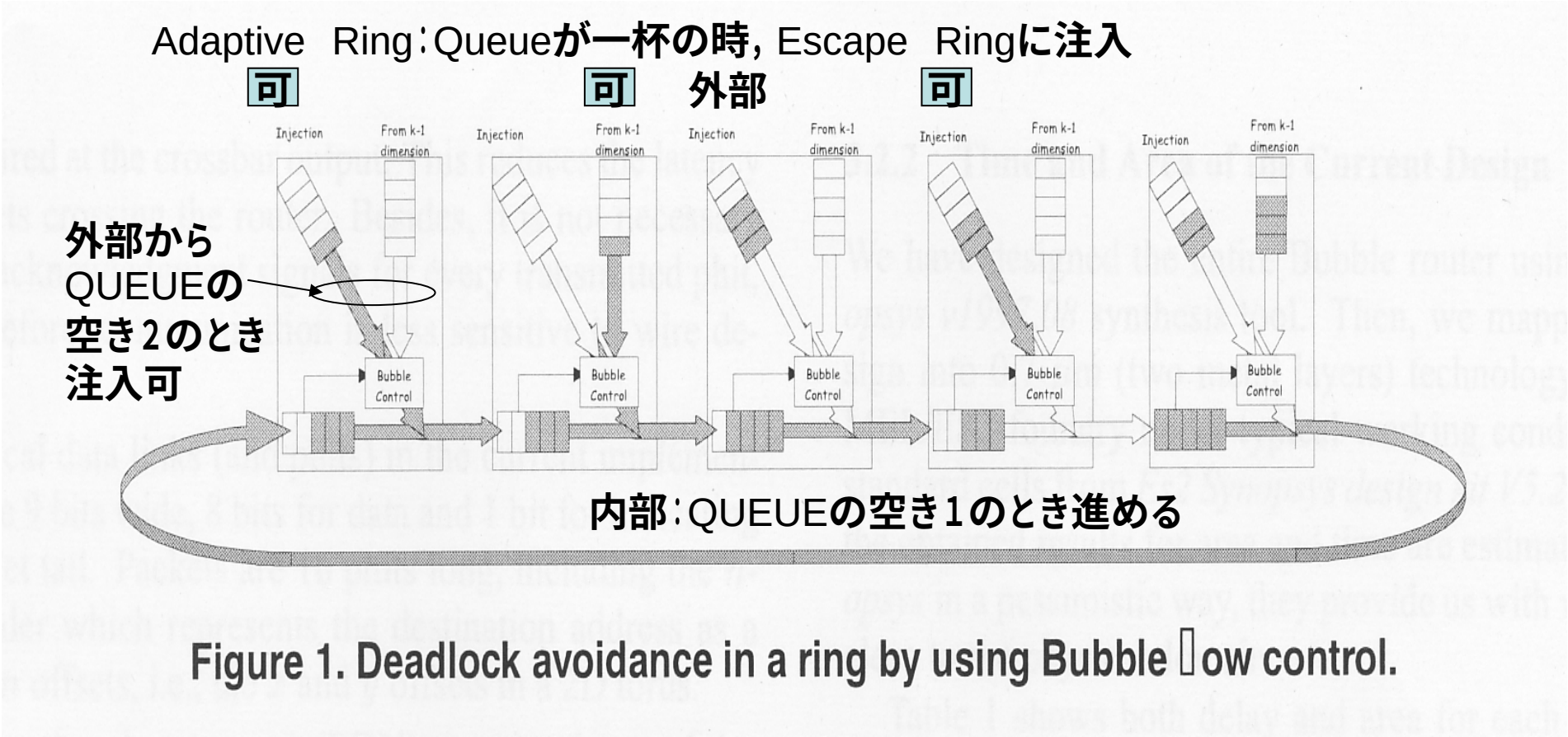
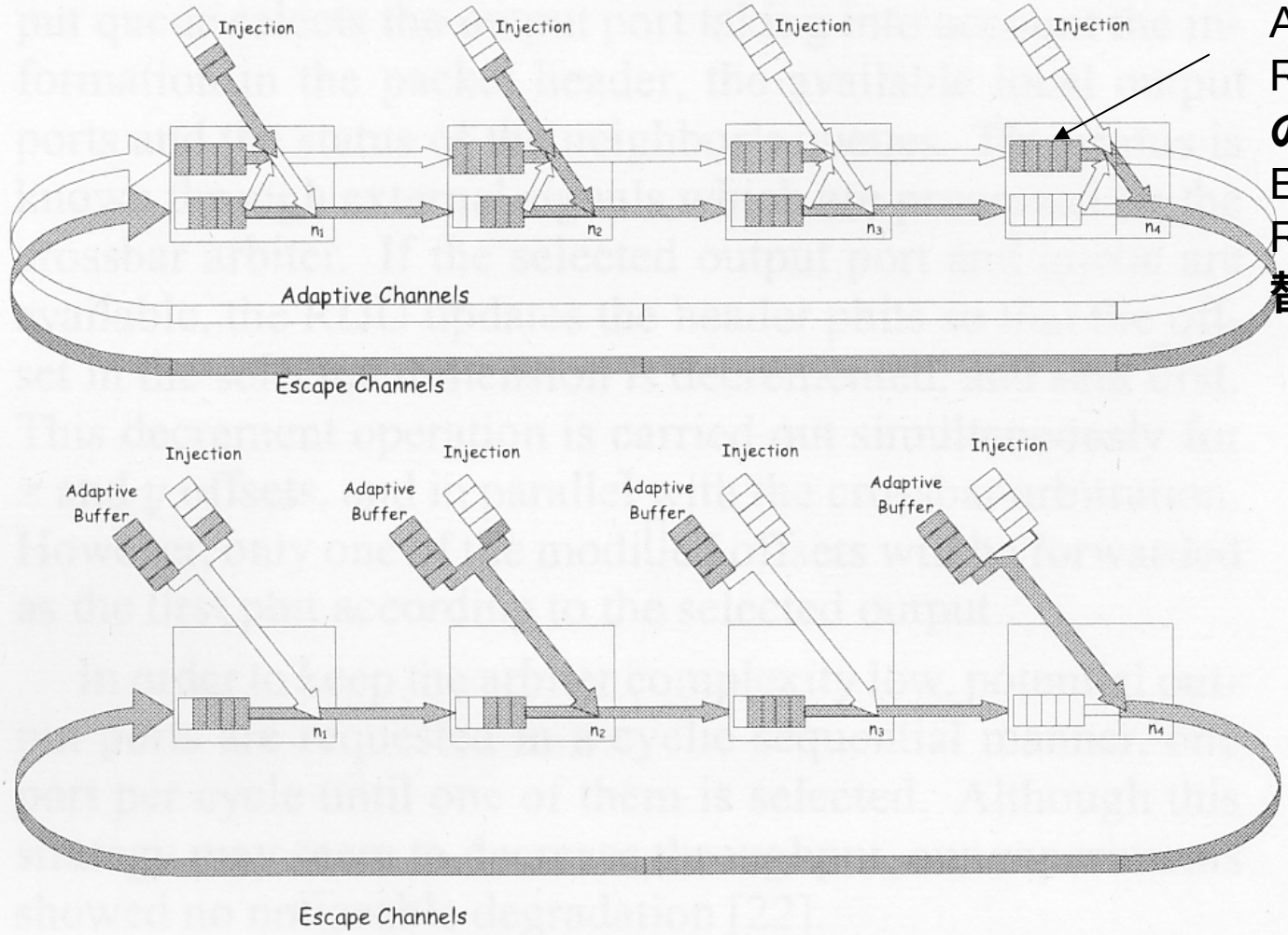


Figure 1. Deadlock avoidance in a ring by using Bubble □ow control.

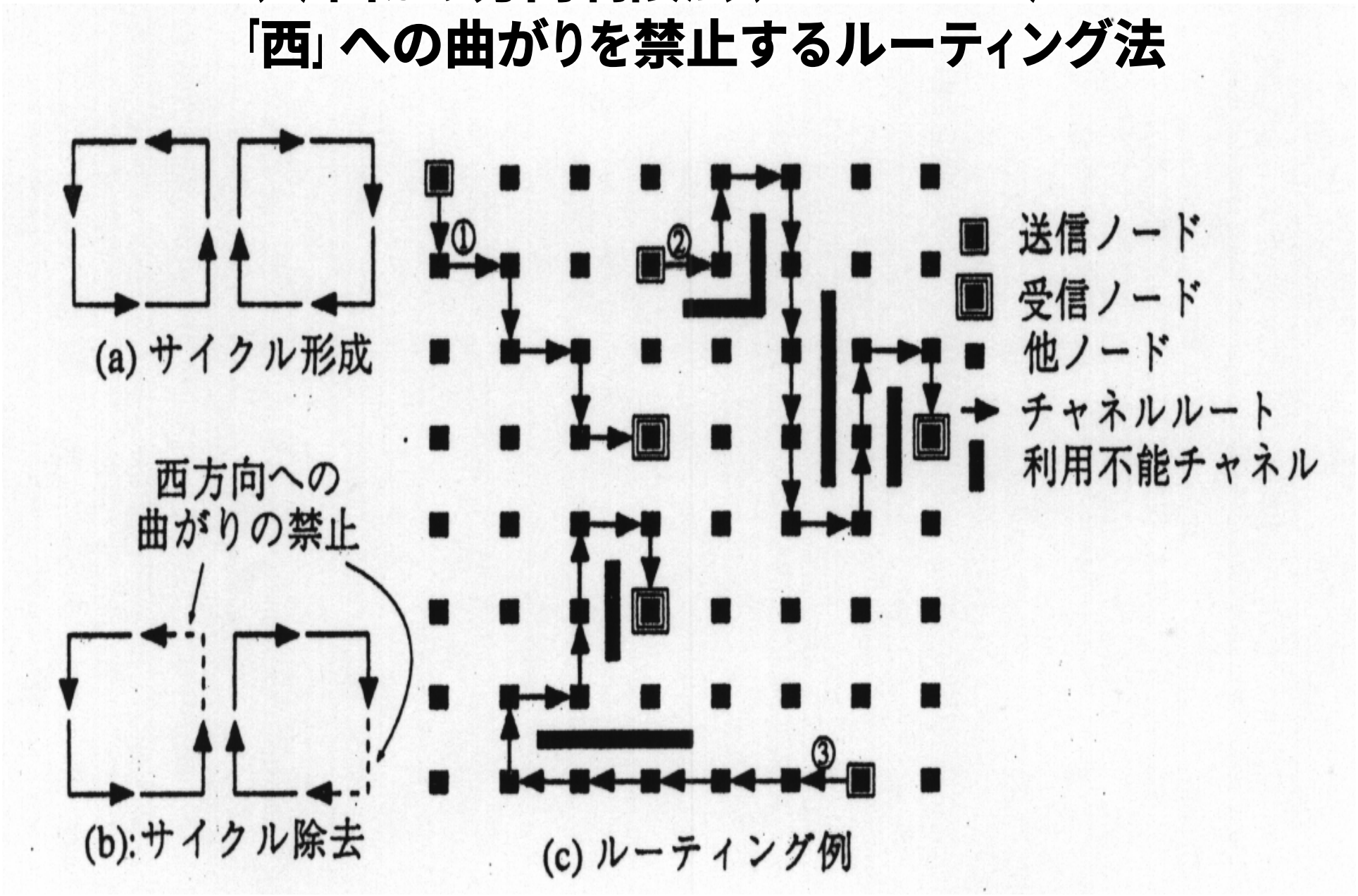


Adaptive Ringがフルのとき、Escape Ringへ切り替え

Figure 2. Representation of adaptive and escape queues when all adaptive ones are busy and equivalent network.

(4) 曲がり方向制限法 (turn model)

「西」への曲がりを禁止するルーティング法



２．４動的網

２.４.１クロスバ網

①集中制御方式

②分散制御方式

２.４.２多段結合網

（１）完全非閉塞網

通信パターン数：N!

通信変更：ローカルにスイッチ変更で対処可

①クロスバ網

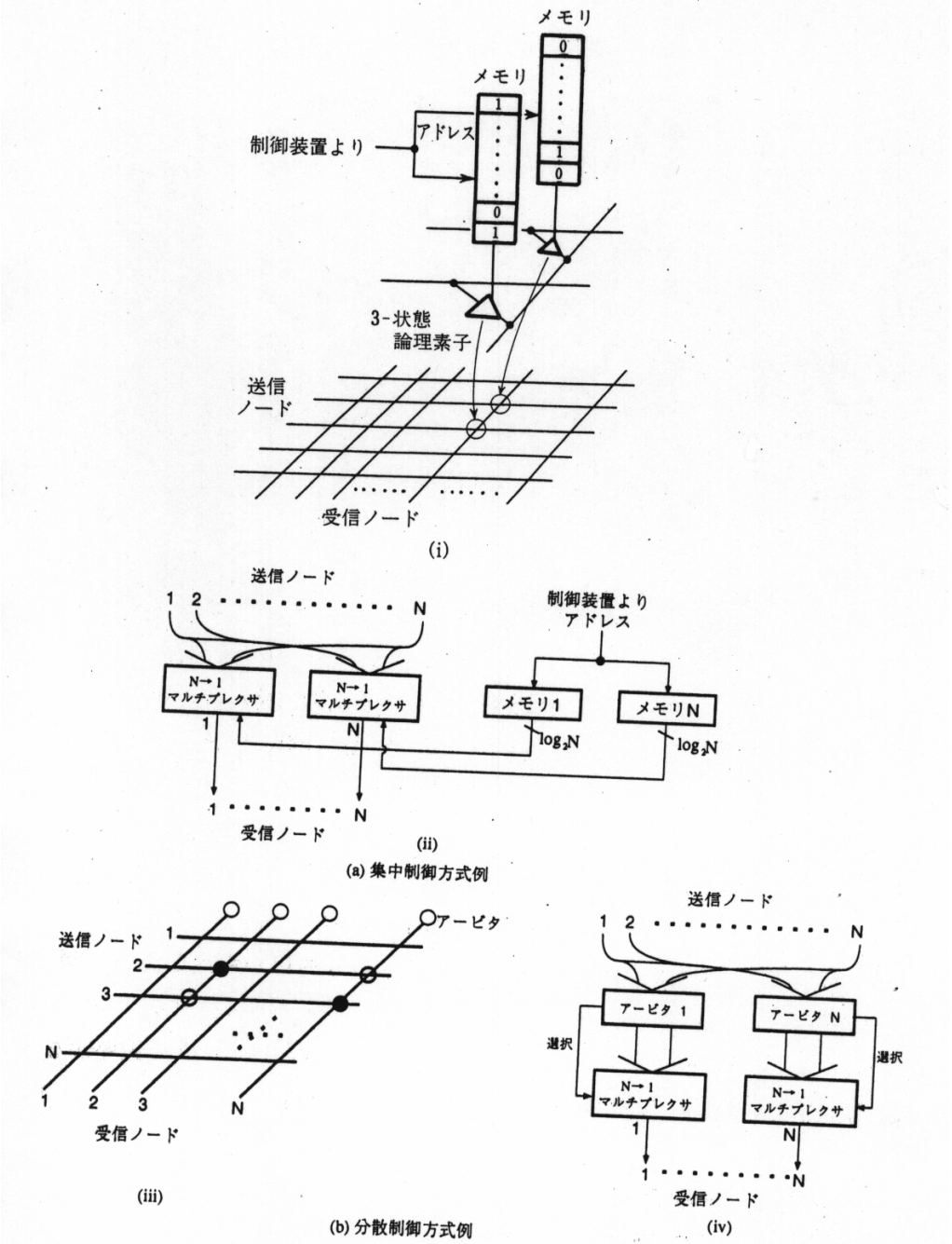

メモリ
メモリ
制御装置より
アドレス
3-状態
論理素子
送信
ノード
受信ノード
(i)
送信ノード
制御装置より
アドレス
N→1
マルチプレクサ
N→1
マルチプレクサ
メモリ1
メモリN
log2N
log2N
受信ノード
(ii)
(a) 集中制御方式例
送信ノード
アービタ
受信ノード
(iii)
送信ノード
アービタ 1
アービタ N
選択
選択
N→1
マルチプレクサ
N→1
マルチプレクサ
受信ノード
(iv)
(b) 分散制御方式例

図 2.20　クロスバ網

②３ステージClos網

m≧2n-1のとき非閉塞：存在定理

スイッチ数

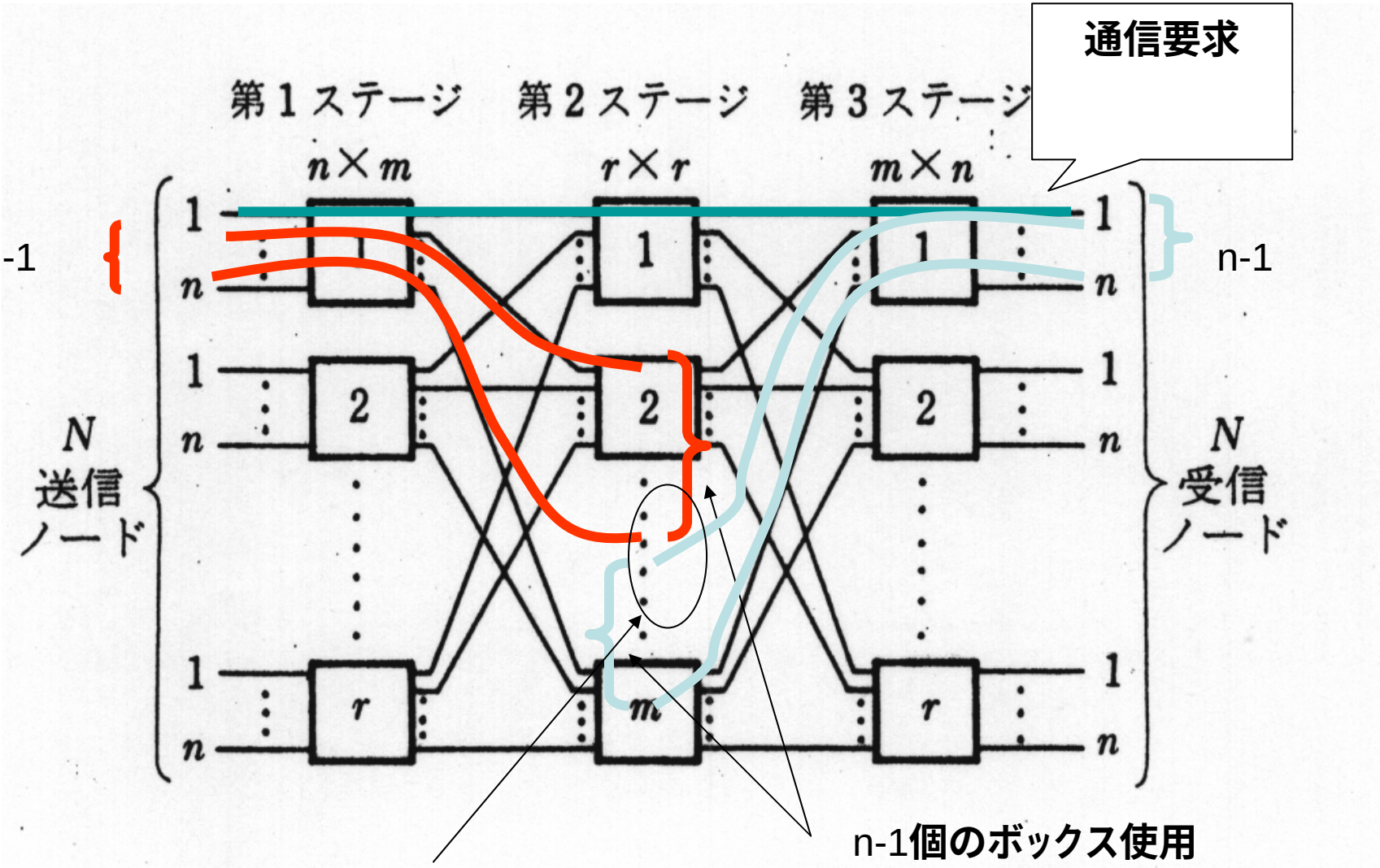
$$Ns=(2n-1)\{(N/n)^2+2N\}$$

$$MinNs=4\sqrt{2}N^{3/2}-4N$$

Ｎ＝２５６の時

22146個　　クロスバ：65536個

ルーティング困難


通信要求
第１ステージ
第２ステージ
第３ステージ
n×m
r×r
m×n
n-1
n-1
N
送信
ノード
N
受信
ノード
n-1個のボックス使用
重なりなし！

（２）再構成型非閉塞網

通信パターン数：N!

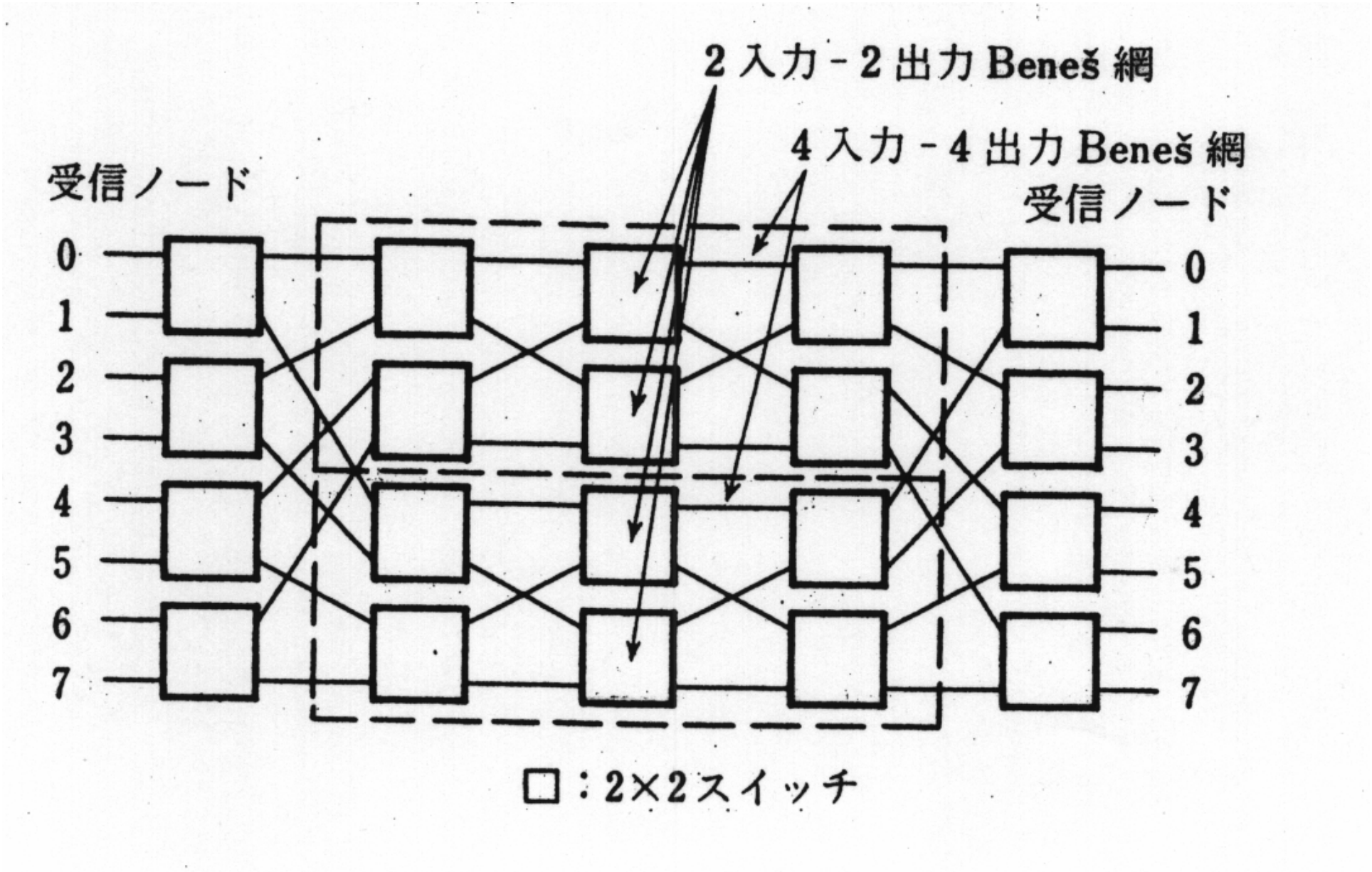
通信変更：ローカルにスイッチ変更で対処不可

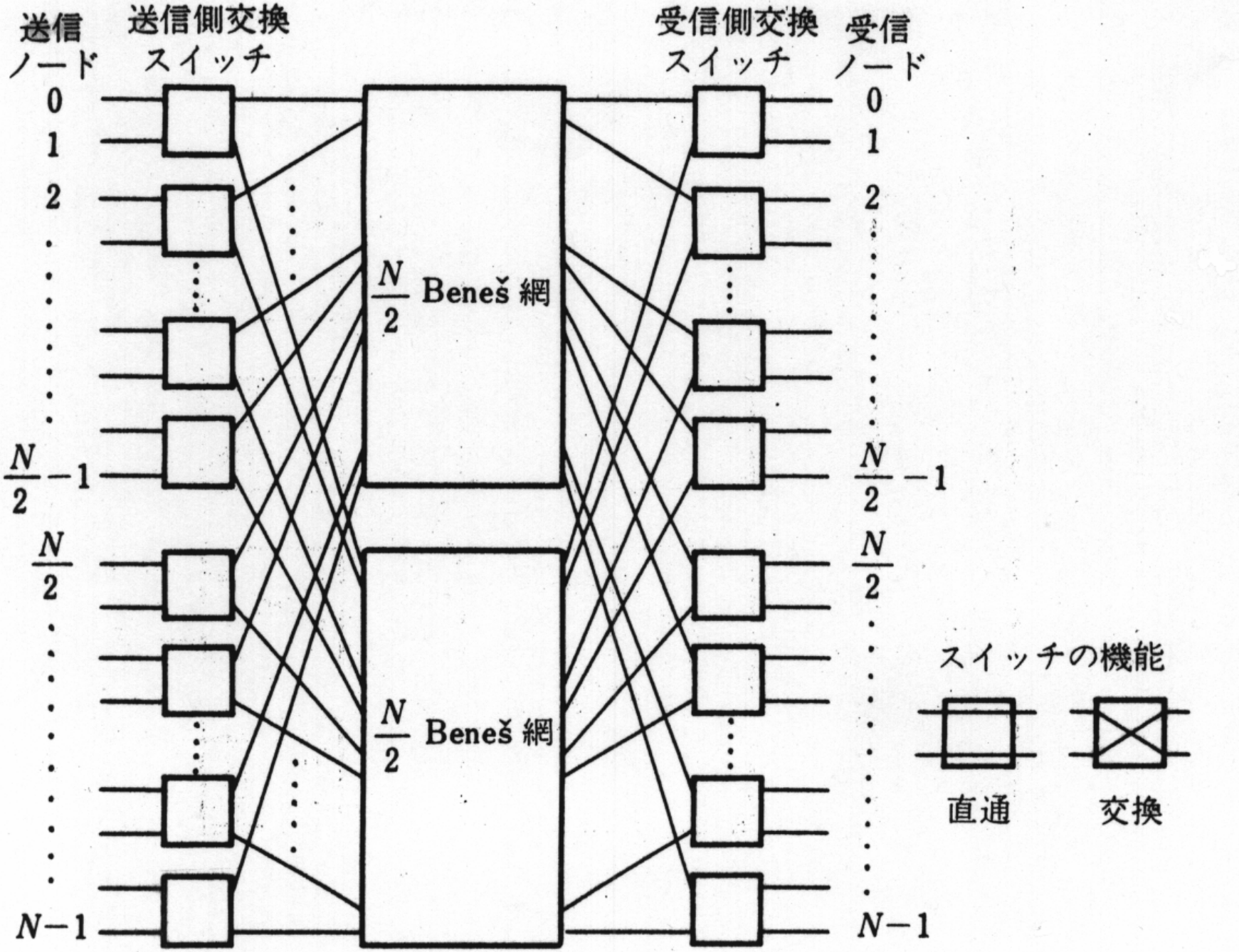
①Benes網

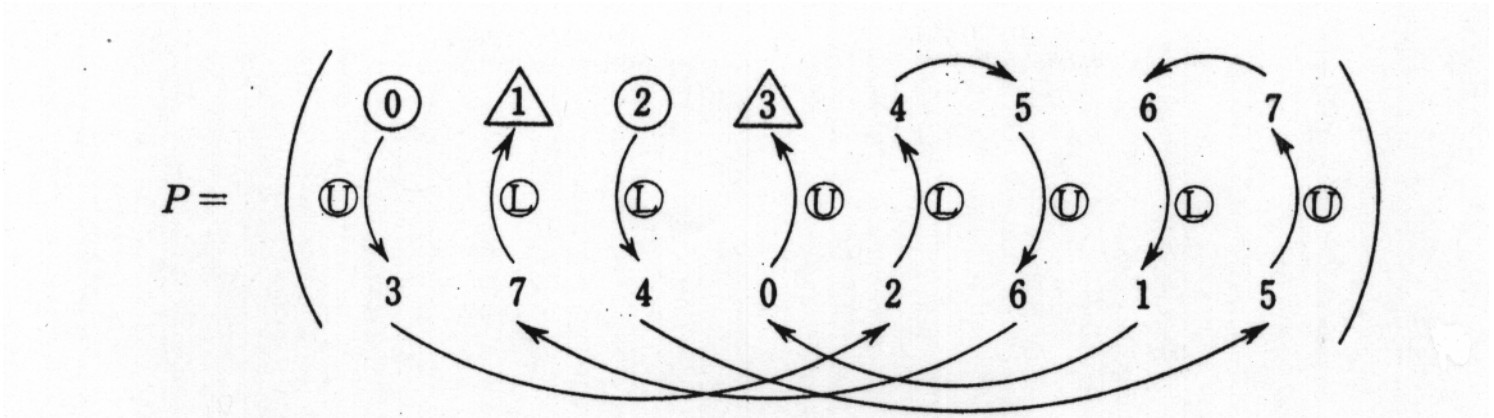
送信ノード　０　１　２　３　４　５　６　７

受信ノード　３　７　４　０　２　６　１　５

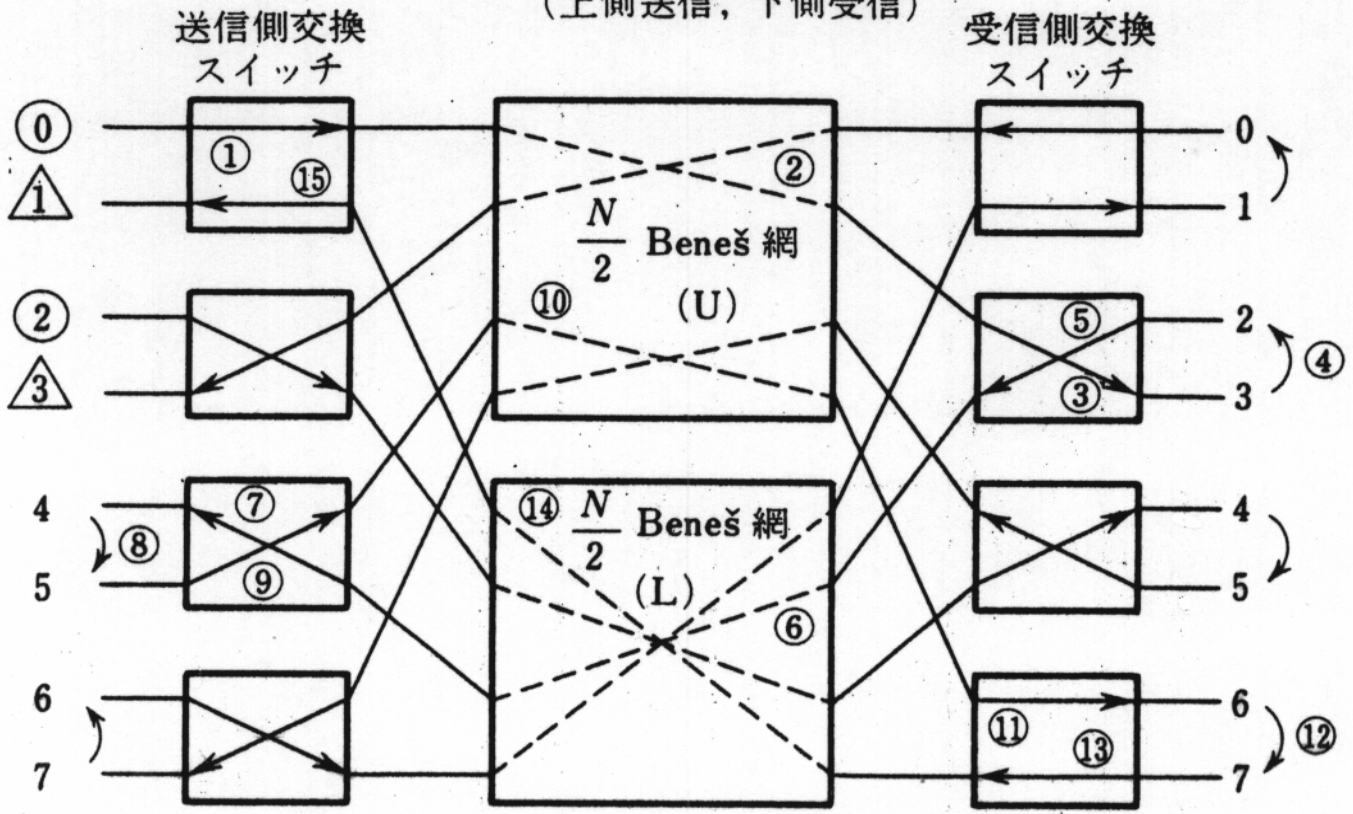
経路選択：ルーピングアルゴリズム


2入力-2出力Beneš網
4入力-4出力Beneš網
受信ノード
受信ノード
0
1
2
3
4
5
6
7
0
1
2
3
4
5
6
7
□：2×2スイッチ


送信
ノード
送信側交換
スイッチ
受信側交換
スイッチ
受信
ノード
0
1
2
N/2 −1
N/2
N−1
N/2 Beneš 網
N/2 Beneš 網
スイッチの機能
直通
交換



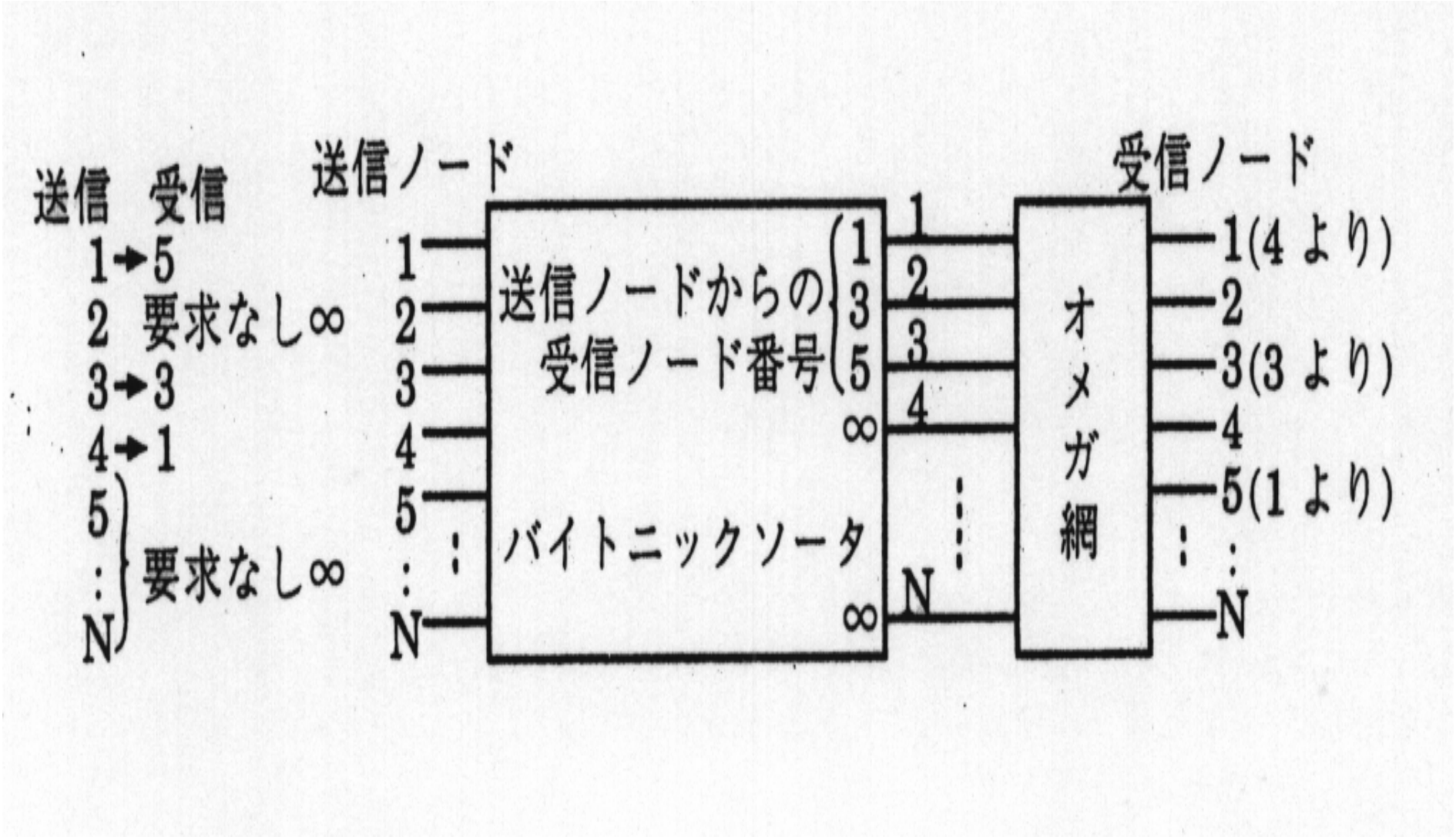
（a） 通信パターン
（上側送信，下側受信）



（b） スイッチ設定例

②バイトニックソータを用いた結合網

バイトニックソータ（図参照）＋オメガ網



非閉塞の証明

$(a_n, a_{n-1}, \ldots, a_2, a_1) > (a^*_n, a^*_{n-1}, \ldots, a^*_2, a^*_1)$

$(b_n, b_{n-1}, \ldots, b_2, b_1) > (b^*_n, b^*_{n-1}, \ldots, b^*_2, b^*_1)$ なら

$(a_i, a_{i-1}, \ldots, a_2, a_1, b_n, b_{n-1}, \ldots, b_{i+1}) \neq$

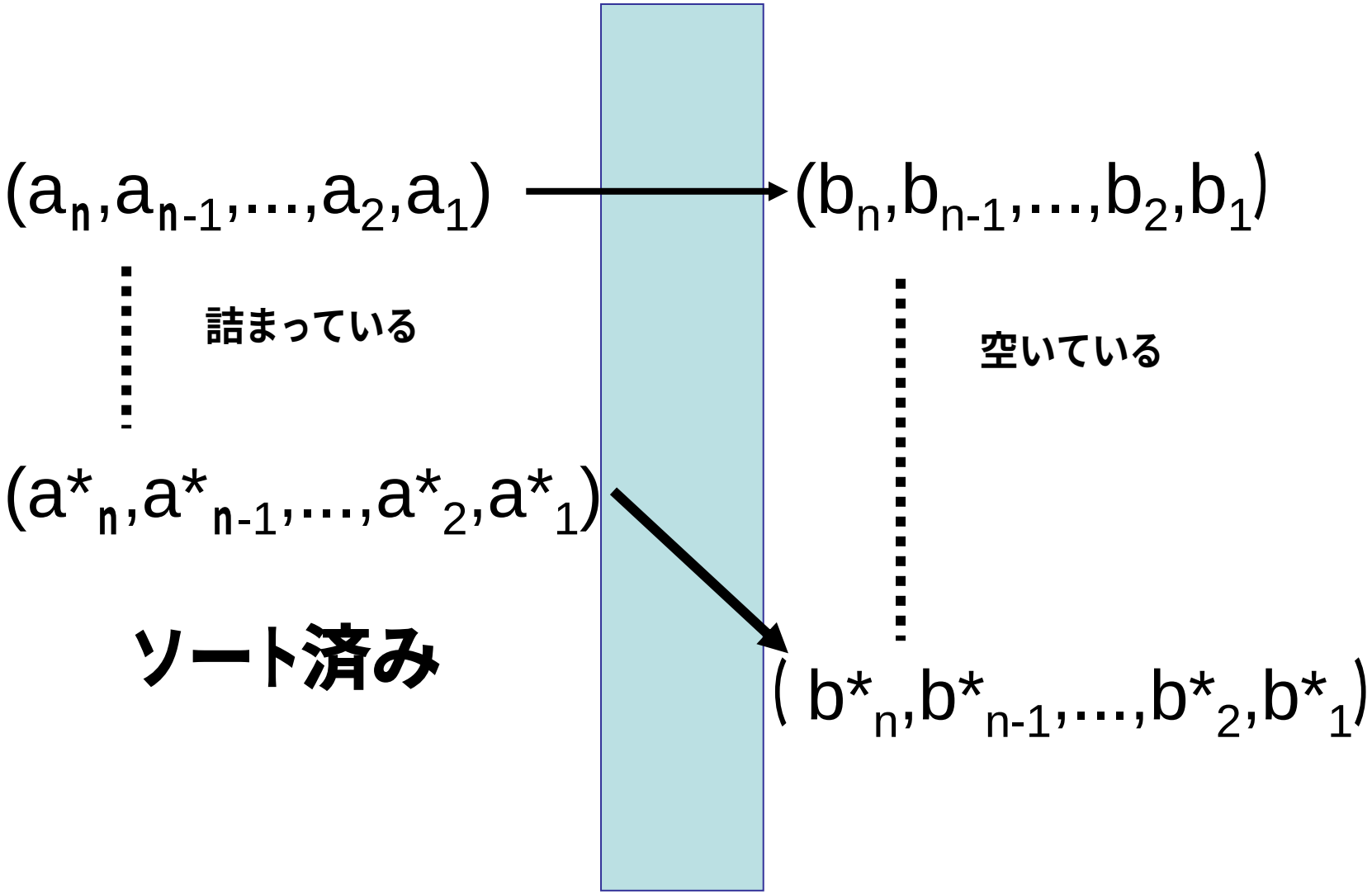
$(a^*_i, a^*_{i-1}, \ldots, a^*_2, a^*_1, b^*_n, b^*_{n-1}, \ldots, b^*_{i+1})$ の証明

$(a_i, a_{i-1}, \ldots, a_2, a_1) = (a^*_i, a^*_{i-1}, \ldots, a^*_2, a^*_1)$ のとき

$(b_n, b_{n-1}, \ldots, b_2, b_1) - (b^*_n, b^*_{n-1}, \ldots, b^*_2, b^*_1) \geqq$

$(a_n, a_{n-1}, \ldots, a_2, a_1) - (a^*_n, a^*_{n-1}, \ldots, a^*_2, a^*_1) \geqq 2^i$ なので

$(b_n, b_{n-1}, \ldots, b_{i+1}) > (b^*_n, b^*_{n-1}, \ldots, b^*_{i+1})$


$(a_n,a_{n-1},\ldots,a_2,a_1)$
$(b_n,b_{n-1},\ldots,b_2,b_1)$
詰まっている
空いている
$(a^*_n,a^*_{n-1},\ldots,a^*_2,a^*_1)$
ソート済み
$(b^*_n,b^*_{n-1},\ldots,b^*_2,b^*_1)$
オメガ網

③３ステージClos網

$2n-1 > m \geqq n$ ：再構成型非閉塞網

Hallの補助定理

A：集合

A_1，A_2，A_3,...，A_r：部分集合

「$a_i \in A_i$，$a_j \in A_j$（$i,j=1,2,...,r$）で$a_i \neq a_j$なる

Aの要素がA_i, A_jに存在する」⇔

「任意のｋ個の和集合$\cup_k A_i$に少なくともｋ個の異

なる要素が存在する」

Slepian　Duguidの定理

入力ボックスI_iから繋がっている

　出力ボックスの集合K_i

K_iは出力ボックスOの部分集合

任意のｋ個の部分集合の和$\cup K_i$の異なる

　　要素の数ｔ

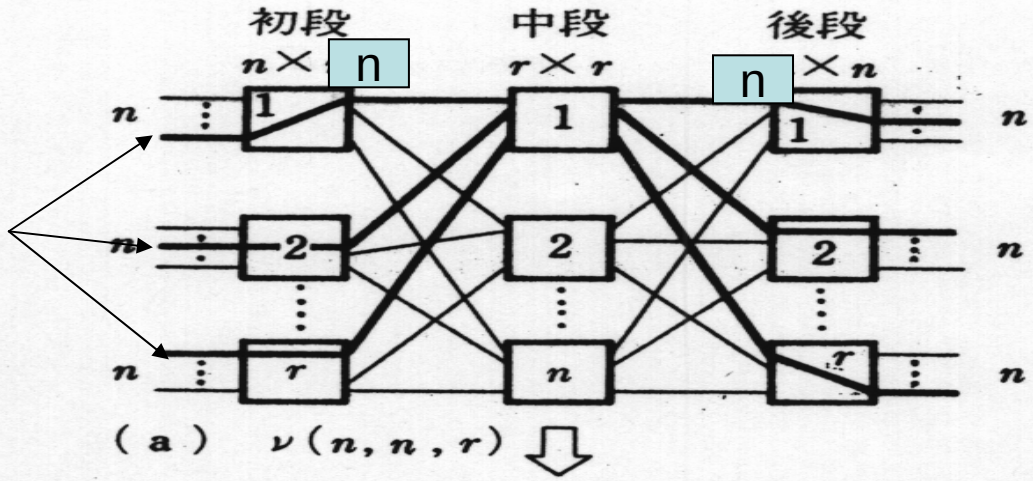
ｎ個の入力ボックスへの入力線数　kn≦

$\cup K_i$の出力ボックスからの出力線数tn

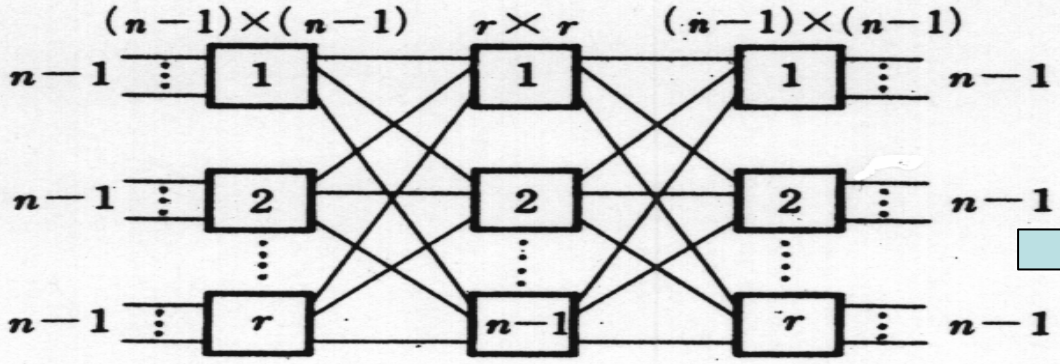
よりｋ≦ｔ．　Hallの補助定理より

　各K_iには異なる要素が存在する

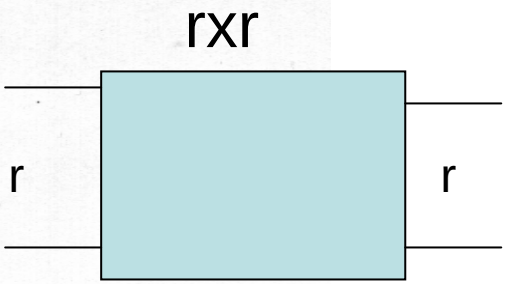
R本の
リンク



（a）　$\nu(n, n, r)$

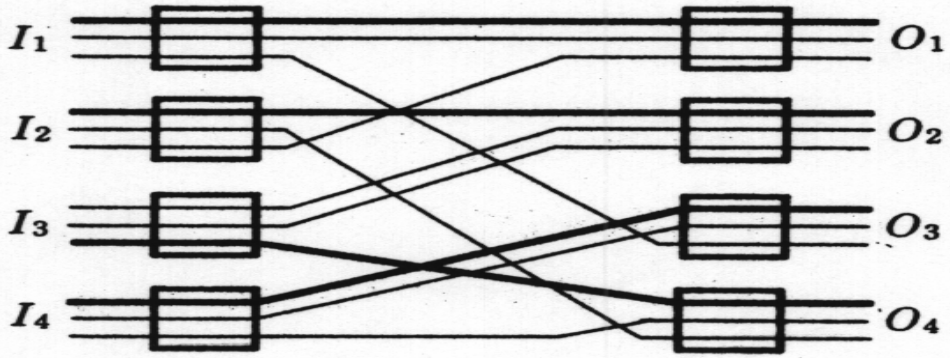


（b）　$\nu(n-1, n-1, r)$



与えられた通信パターンに対して

初段、後段のスイッチボックスを
1つずつ使用するr本のリンクが
存在



（c）　証明のための例

$K_1 = (①, 3)$
$K_2 = (1, ②, 4)$
$K_3 = (2, ④)$
$K_4 = (③, 4)$

（３）閉塞網

可能な通信パターン総数：Ｎ！の一部

$$\left(\sqrt{N}\right)^{N}=2^{(N\log N)/2}$$

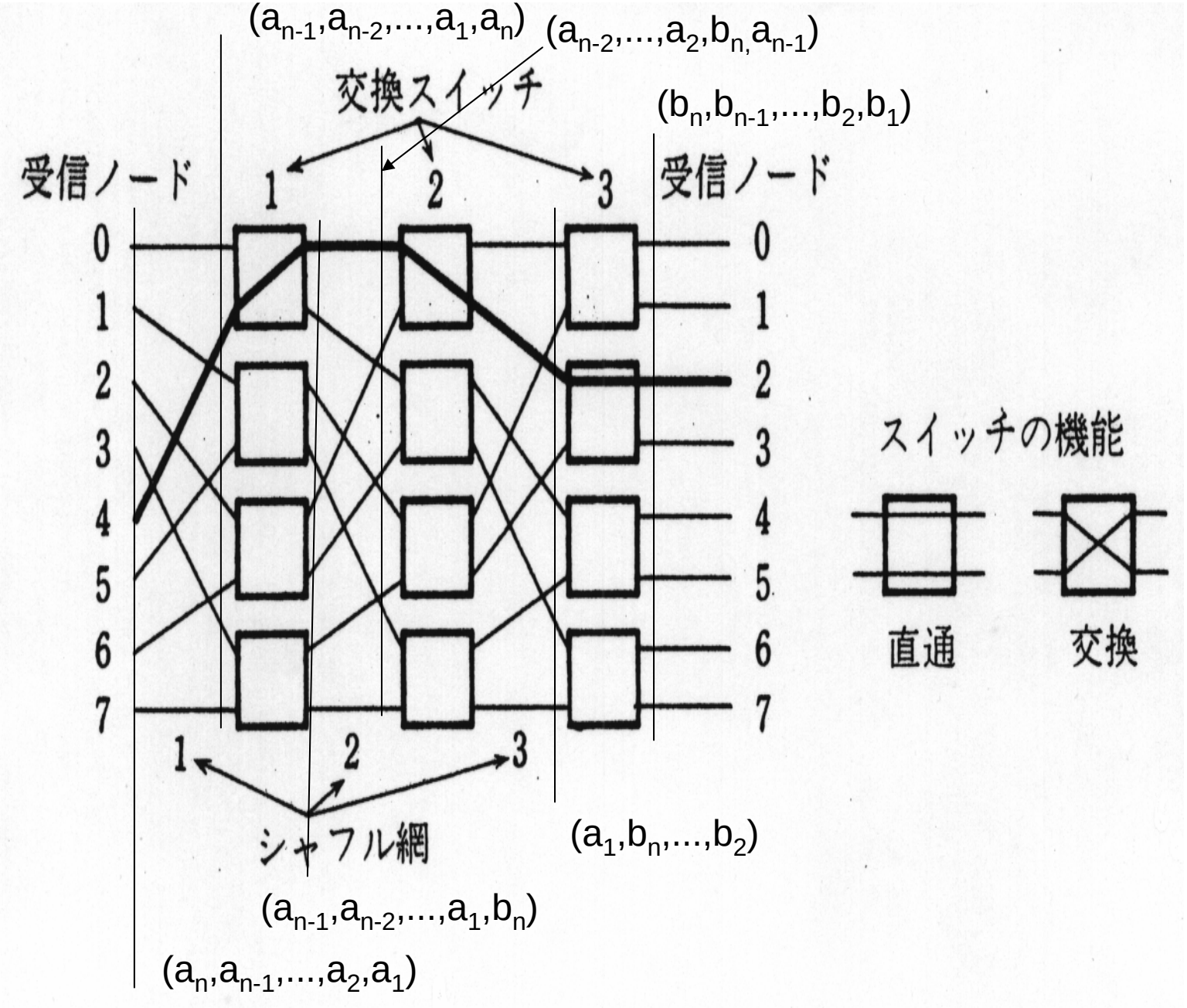
①オメガ網

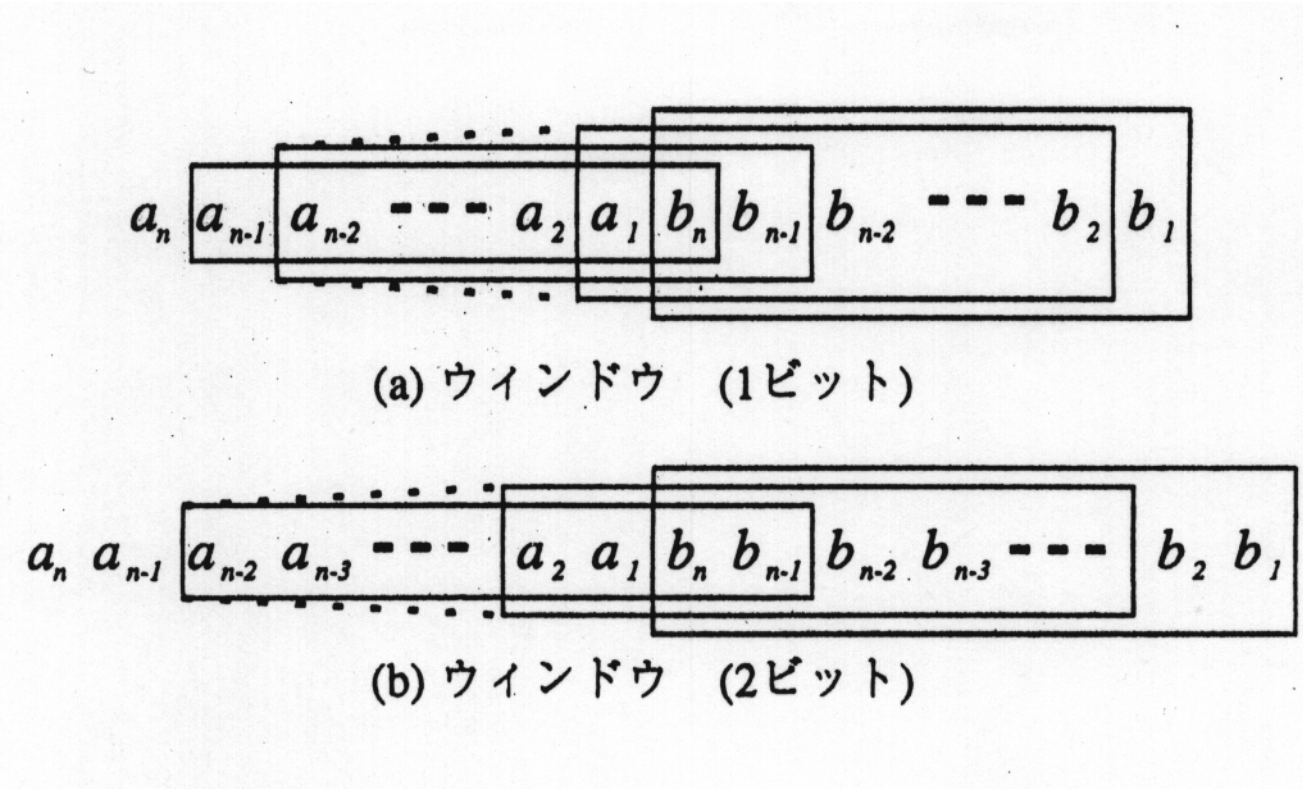
・オメガ網の構成

$$E=(e_1, i)$$

$$\Omega=(\sigma E)^{n}$$

・オメガ網のスイッチ設定


$(a_{n-1},a_{n-2},\ldots,a_1,a_n)$
$(a_{n-2},\ldots,a_2,b_n,a_{n-1})$
交換スイッチ
$(b_n,b_{n-1},\ldots,b_2,b_1)$
受信ノード
1
2
3
受信ノード
0
1
2
3
4
5
6
7
0
1
2
3
4
5
6
7
スイッチの機能
直通
交換
1
2
3
シャフル網
$(a_1,b_n,\ldots,b_2)$
$(a_{n-1},a_{n-2},\ldots,a_1,b_n)$
$(a_n,a_{n-1},\ldots,a_2,a_1)$


a_n a_{n-1} a_{n-2} --- a_2 a_1 b_n b_{n-1} b_{n-2} --- b_2 b_1
(a) ウィンドウ (1ビット)
a_n a_{n-1} a_{n-2} a_{n-3} --- a_2 a_1 b_n b_{n-1} b_{n-2} b_{n-3} --- b_2 b_1
(b) ウィンドウ (2ビット)

送信ノード$(a_n, a_{n-1}, \ldots, a_2, a_1)$

受信ノード$(b_n, b_{n-1}, \ldots, b_2, b_1)$

最初のシャフル：

送信ノード$(a_n, a_{n-1}, \ldots, a_2, a_1)$**は**

$(a_{n-1}, a_{n-2}, \ldots, a_2, a_n)$**に対応**

最初の交換スイッチ：

$a_n \oplus b_n = 0$**であれば、直通側**

$a_n \oplus b_n = 1$**であれば、交換側**

最初の交換スイッチの出力

送信ノード$(a_n, a_{n-1}, \ldots, a_2, a_1)$**は**

$(a_{n-1}, a_{n-2}, \ldots, a_2, b_n)$

第２段のシャフル

$(a_{n-2}, a_{n-3}, \ldots, a_2, b_n, a_{n-1})$

第２段の交換スイッチ：

$a_{n-1} \oplus b_{n-1} = 0$**のとき、直通**

$a_{n-1} \oplus b_{n-1} = 1$**のとき、交換**

第２段交換スイッチの出力

送信ノード$(a_n, a_{n-1}, \ldots, a_2, a_1)$

$(a_{n-2}, a_{n-3}, \ldots, a_2, b_n, b_{n-1})$**に対応**

交換スイッチ

b_i=0**なら、上方リンク選択**

b_i=1**なら、下方リンク選択**

ウィンドウ表現

BxB**スイッチの利用**

$B=2^b$

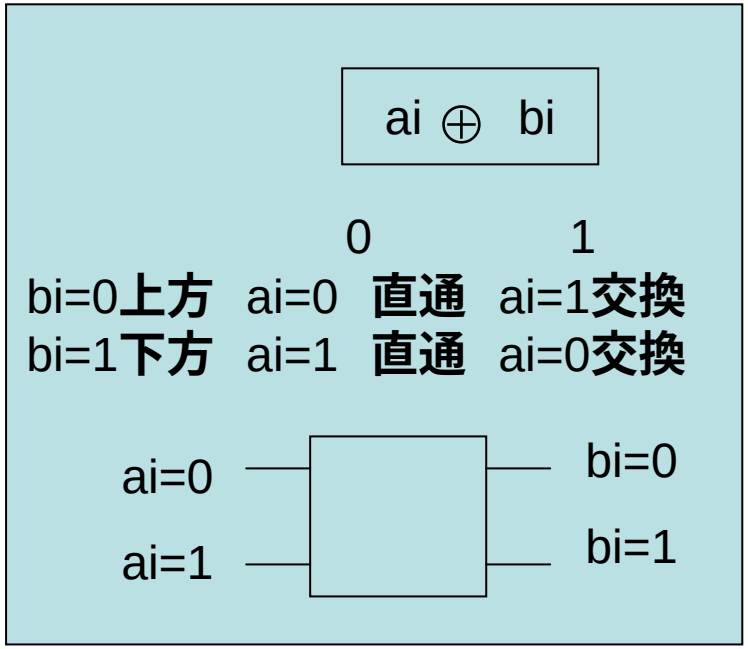
$N=B^k$

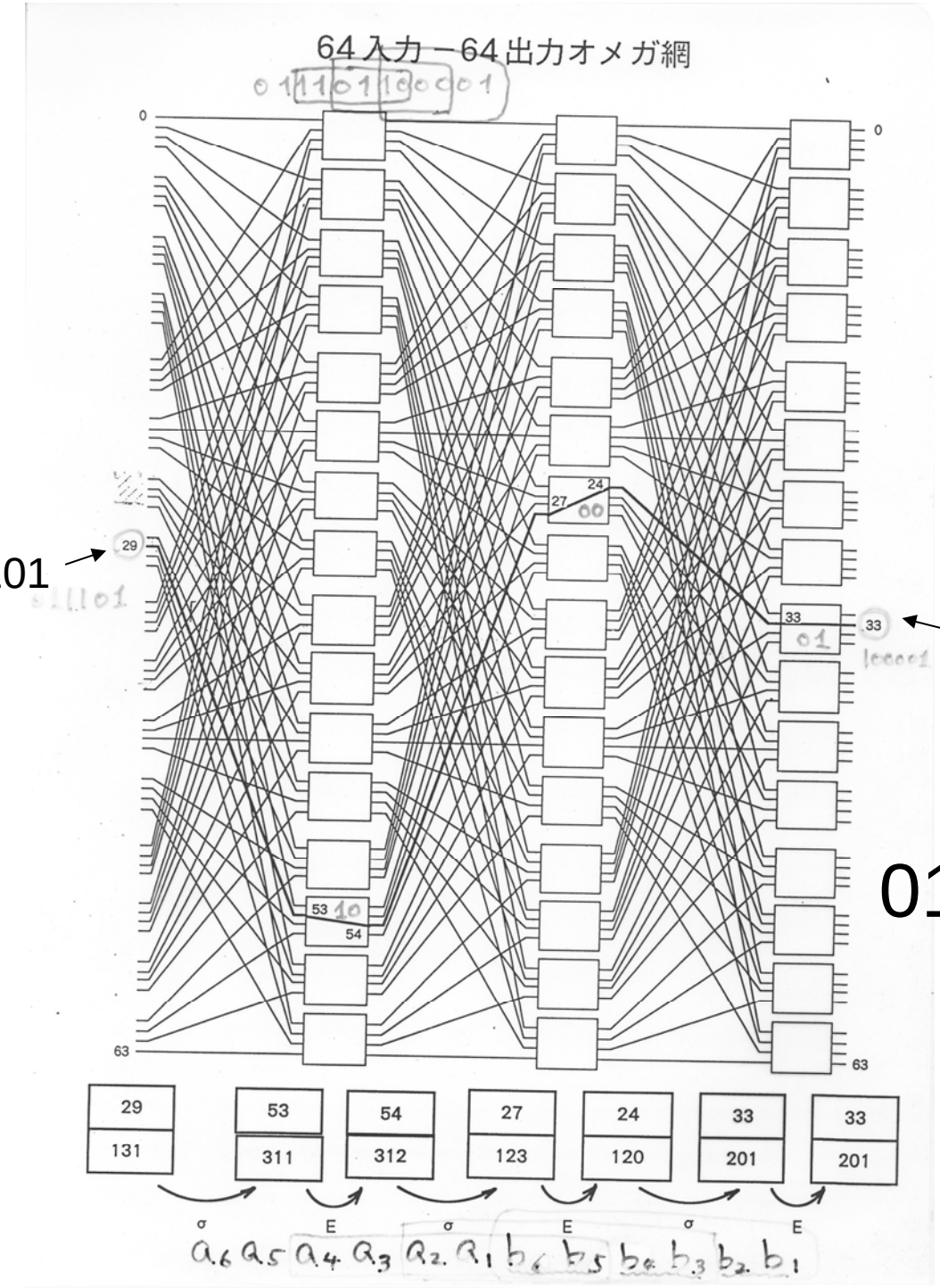
BxB**のスイッチｋ段のオメガ網**

２ビットシャフル、

$(a_n, a_{n-1}, \ldots, a_2, a_1) \rightarrow$

$(a_{n-2}, a_{n-3}, \ldots, a_2, a_1, a_n, a_{n-1})$





011101

100001

011101 100001

②間接２進ｎ－キューブ

$C=E\beta_2E\beta_3,\ldots,E\beta_nE\sigma^{-1}$

③バンヤン（Banyan）網

$Y=E\beta_2E\beta_3,\ldots,E\beta_nE$

④ベースライン（baseline）網

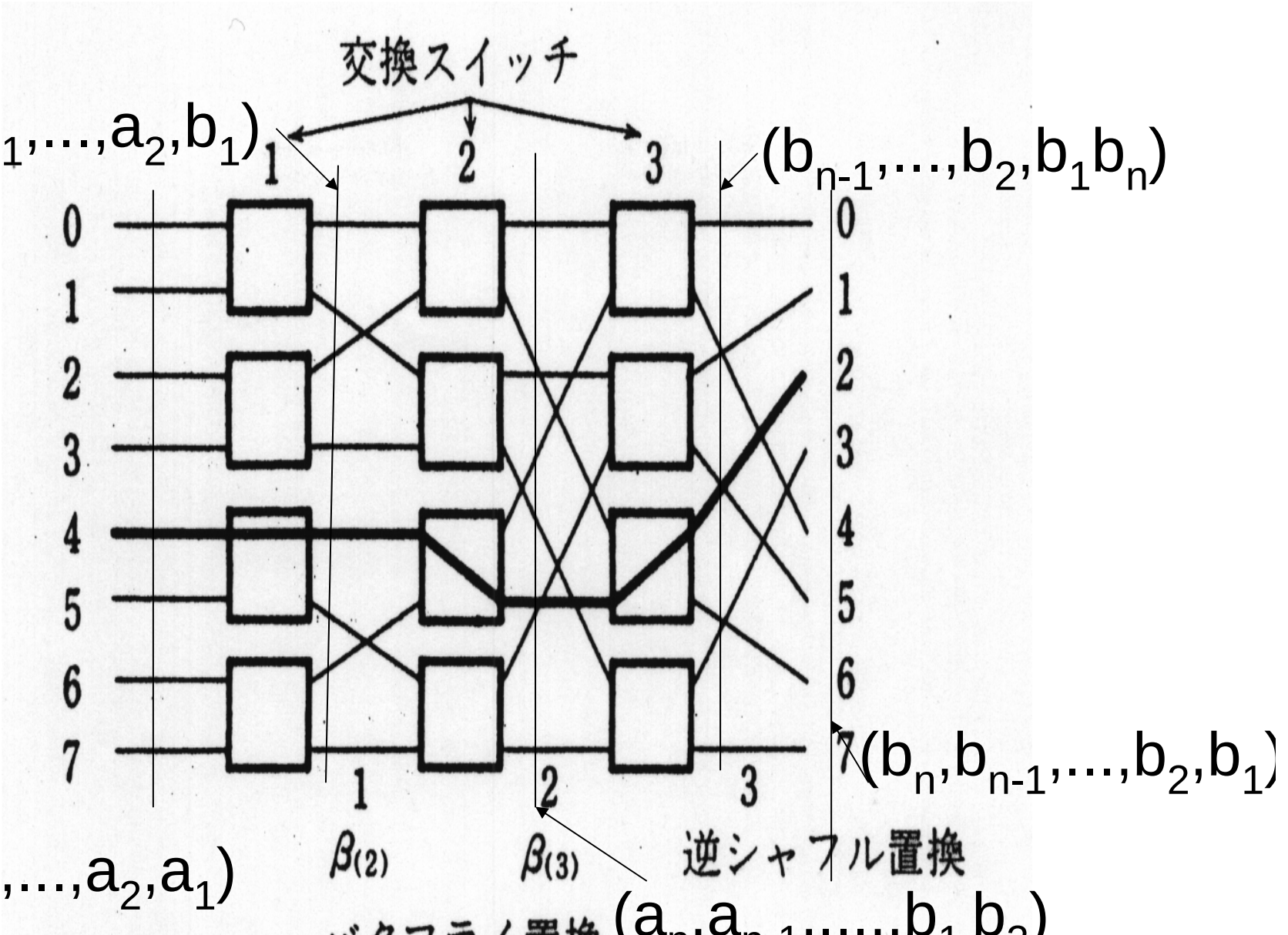
$B=E\sigma_n{}^{-1}E\sigma_{n-1}{}^{-1},\ldots,E\sigma_2{}^{-1}E$

⑤閉塞網間の関係

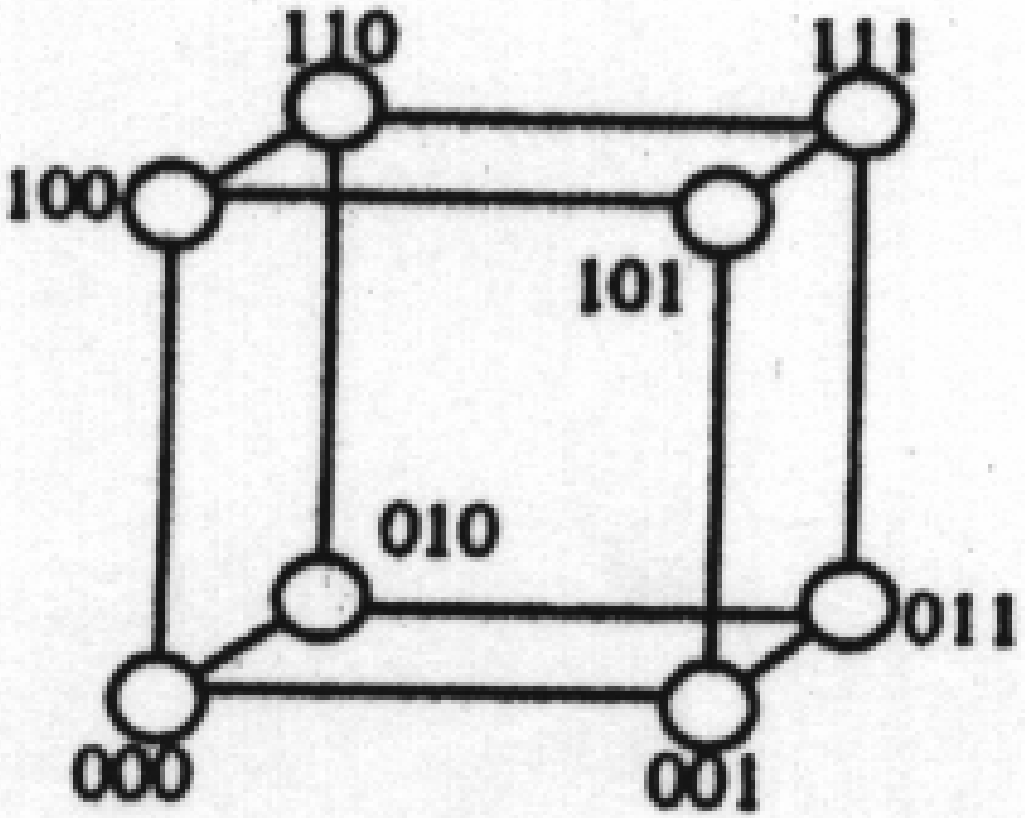
オメガ網（Ω）、間接２進ｎキューブ網（Ｃ），

バンヤン網（Ｙ）

$\Omega^{-1}=C=Y\sigma$


交換スイッチ
$(a_n,a_{n-1},\ldots,a_2,b_1)$
$(b_{n-1},\ldots,b_2,b_1b_n)$
$(b_n,b_{n-1},\ldots,b_2,b_1)$
$(a_n,a_{n-1},\ldots,a_2,a_1)$
$(a_n,a_{n-1},\ldots,,b_1,b_2)$
$\beta_{(2)}$
$\beta_{(3)}$
逆シャフル置換
バタフライ置換

間接2進n-キューブ



(a)2進 3-キューブ

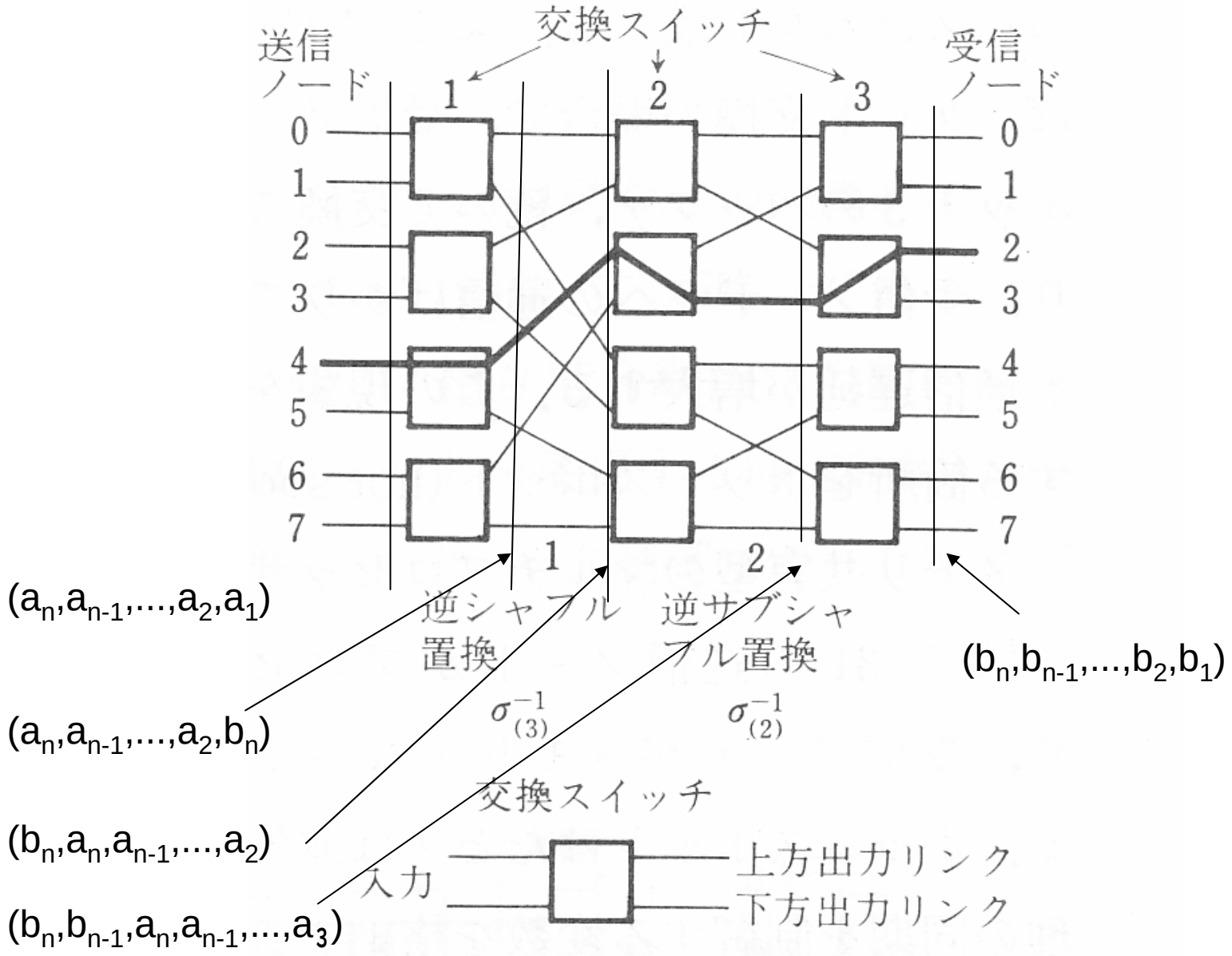
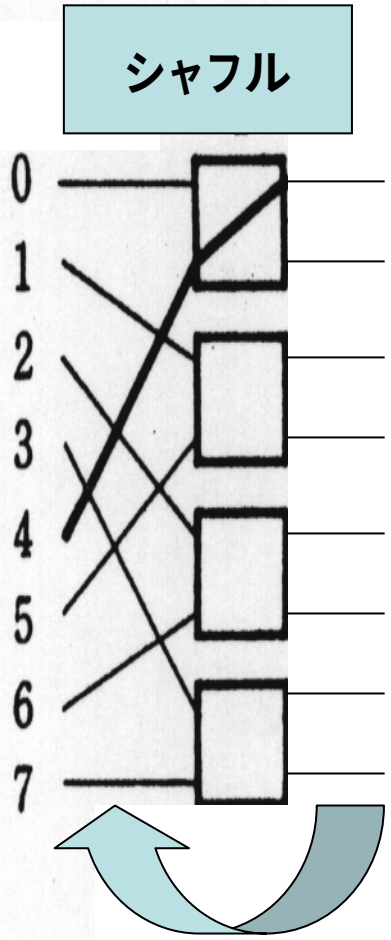


図 2.30　ベースライン網

単一段結合網



単一段シャフルネットワーク

２．５多段結合網の通信制御と耐故障設計

2.5.1木飽和と結合操作

例　ベクトル処理

プロセッサ$P_0,..P_7$：

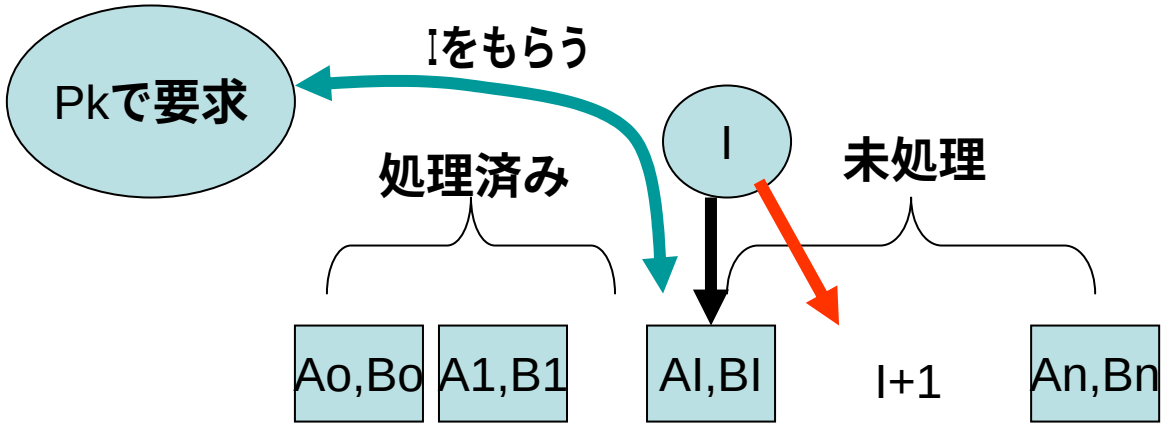
A(I)＋B(I)

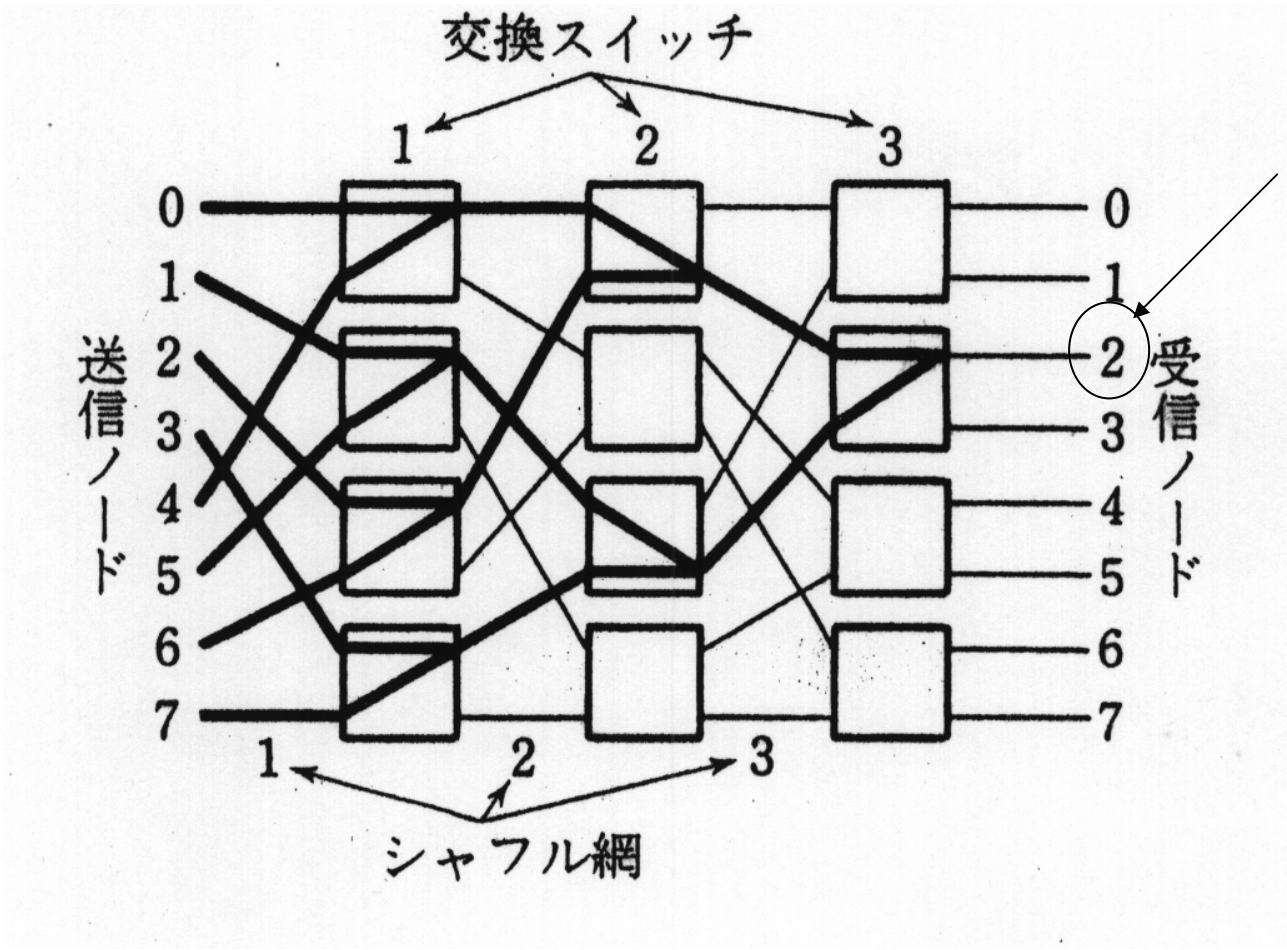
Iを獲得し，＋１

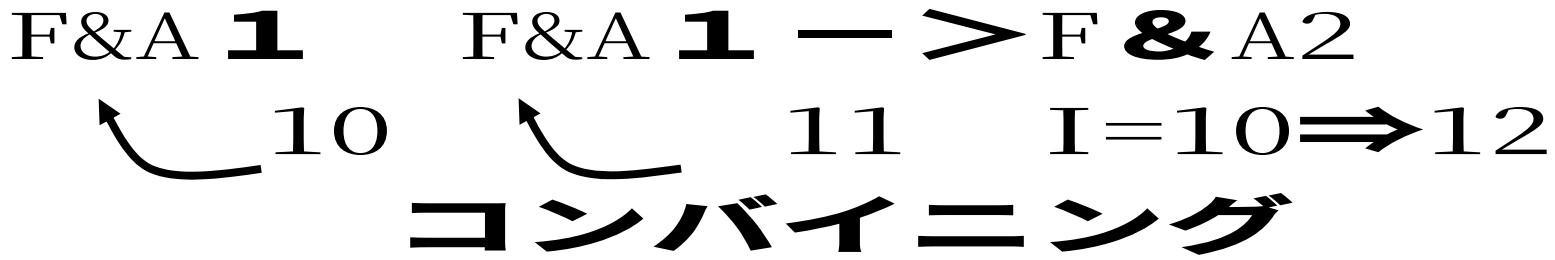
共有変数I：ホットスポット

Fetch and Add 命令

あるスイッチで２つの要求




交換スイッチ
1
2
3
0
1
2
3
4
5
6
7
送信ノード
受信ノード
制御変数格納
シャフル網



コンバイニング

２．５．２耐故障網

①送信側に付加ステージを追加する方式

②各ステージ内にループ構造を導入する方式

③大きなクロスバスイッチを用いる方式

クロスバスイッチの大きさ：$B \times B (B=2^b)$、

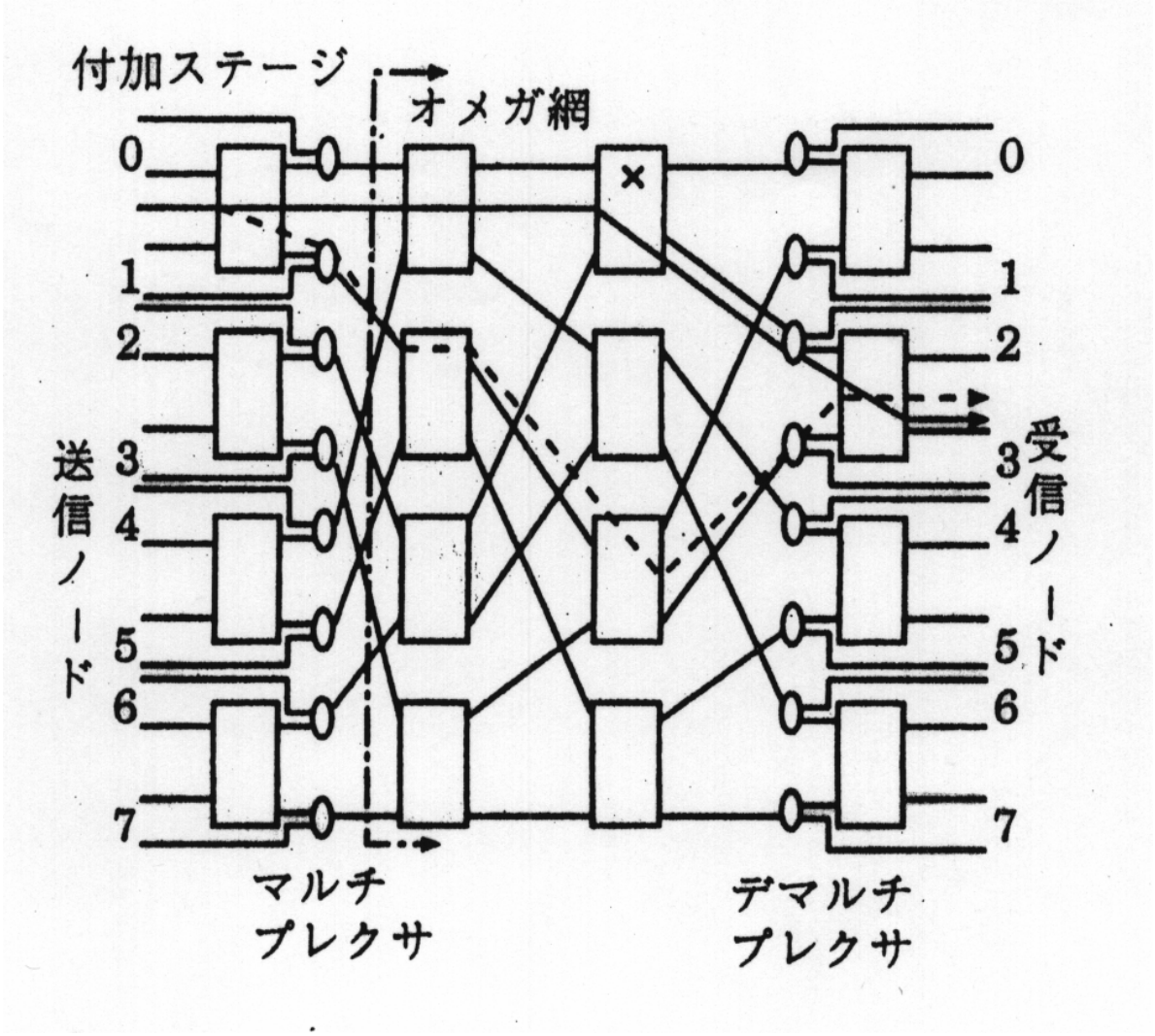
ノード数$N (=2^n)$

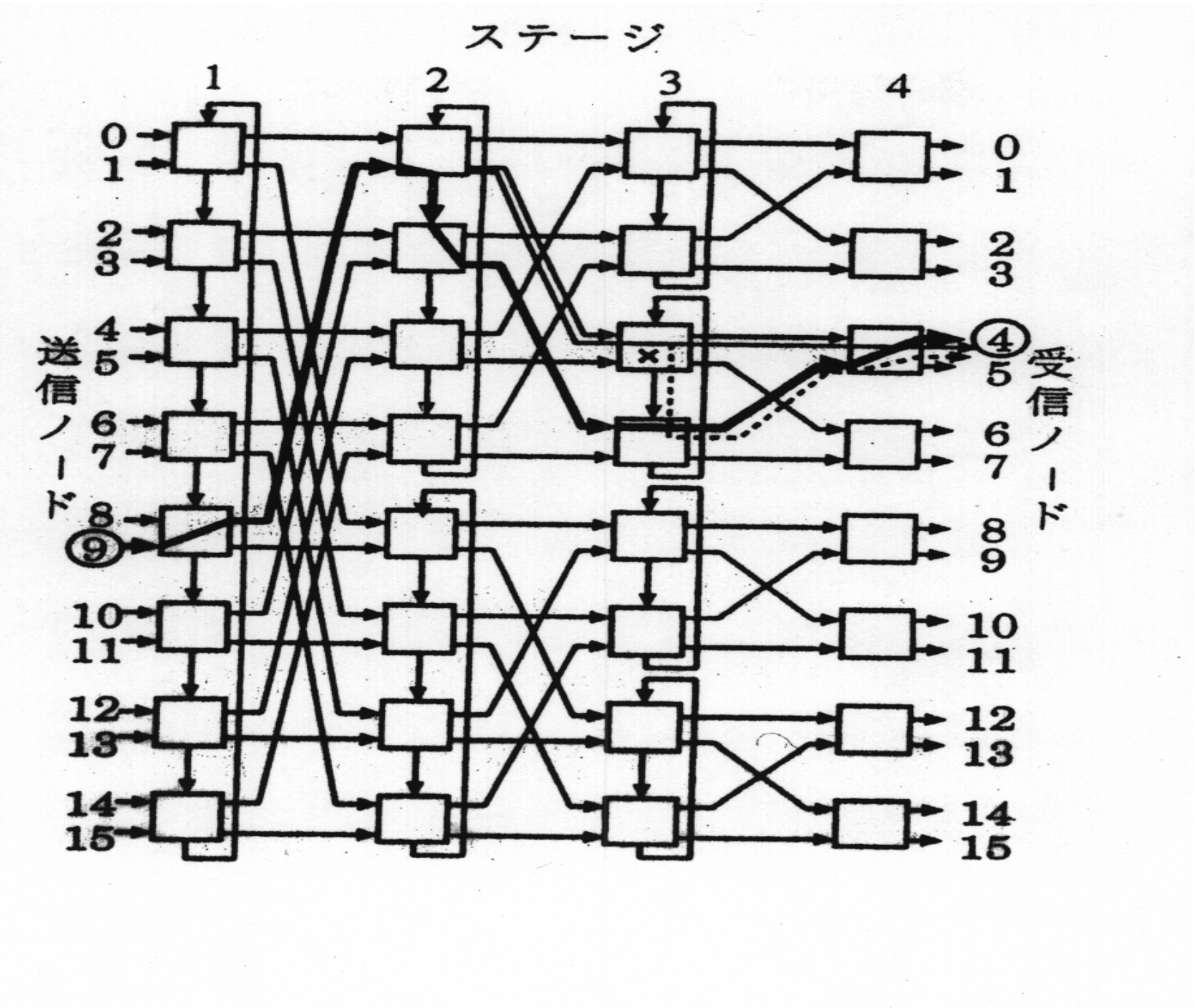
Ｒーパス可能な$\lceil \mathrm{Log}_B N \rceil$段の

マルチパスオメガ網

$R=B^{\lceil n/b \rceil}-n/b$

$N=2^n=B^k$のとき、$R=1$


付加ステージ
オメガ網
0
1
2
3
4
5
6
7
送信ノード
受信ノード
マルチプレクサ
デマルチプレクサ


ステージ
1
2
3
4
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
送信ノード
受信ノード

（１－パスオメガ網）

kの設定

$2^{b(k-1)} < N < 2^{bk}$

$k=\lceil \mathrm{Log}_B N \rceil = \lceil n/b \rceil$

kb-n個の＊

$a_n a_{n-1} \cdots a_{n-b} \cdots a_{n-2b} \cdots a_1 ****b_n b_{n-1} \cdots b_2 b_1$

長さnビットのウィンドウ：bビット毎、k個

$kb-n<b$

最終ステージ以外のウィンドウ：

kb-n個の＊印をすべて含む。

$R=2^{kb-n}=B^{\lceil n/b \rceil - n/b}$

N=32、B=4の３ステージのオメガ網

ノード８から１２への通信

ウィンドウ０１０００＊０１１００

＊＝０の時

各ステージでの受信ノード番号

００００、０００１１、０１１００

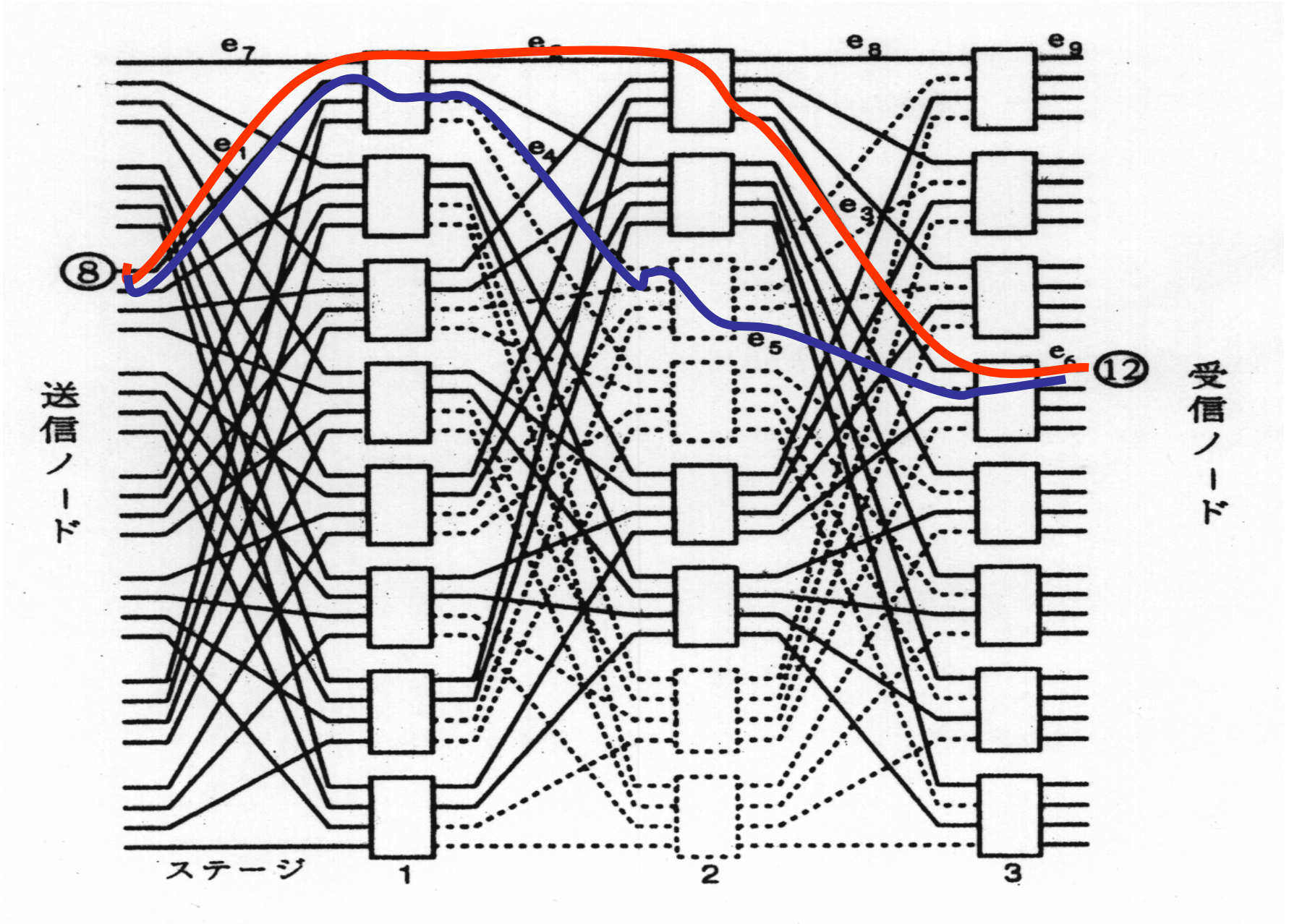
経路e_1、e_2、e_3、e_6

＊＝１の時

各ステージでの受信ノード番号

００００１０、０１０１１、０１１００

経路e_1、e_4、e_5、e_6


e7
e8
e9
e1
e4
e3
e5
e6
⑧
⑫
送信ノード
受信ノード
ステージ
1
2
3

２.５.３負荷分散網

１．　６可変構造型

相互結合網

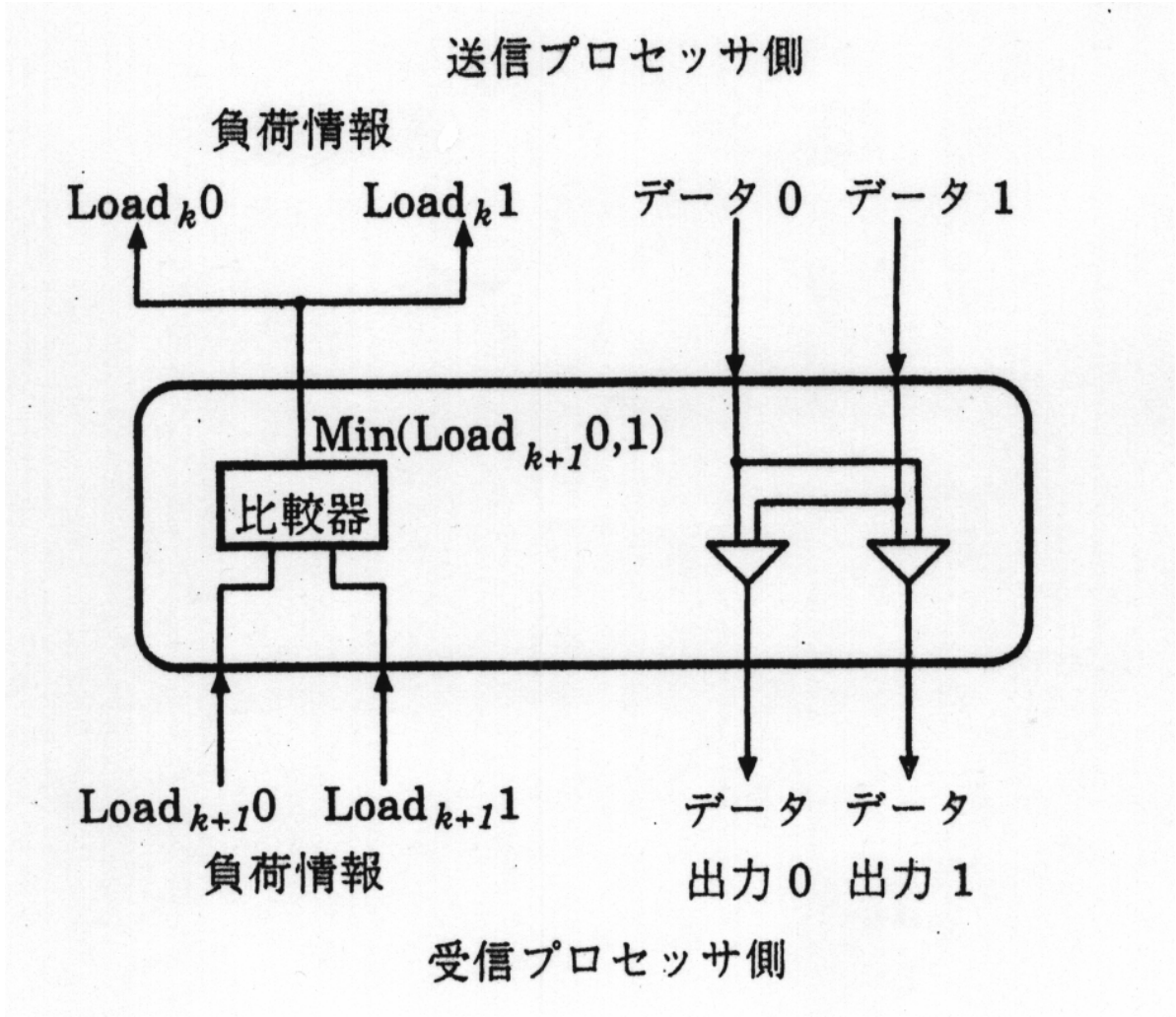
①機械的可変構造

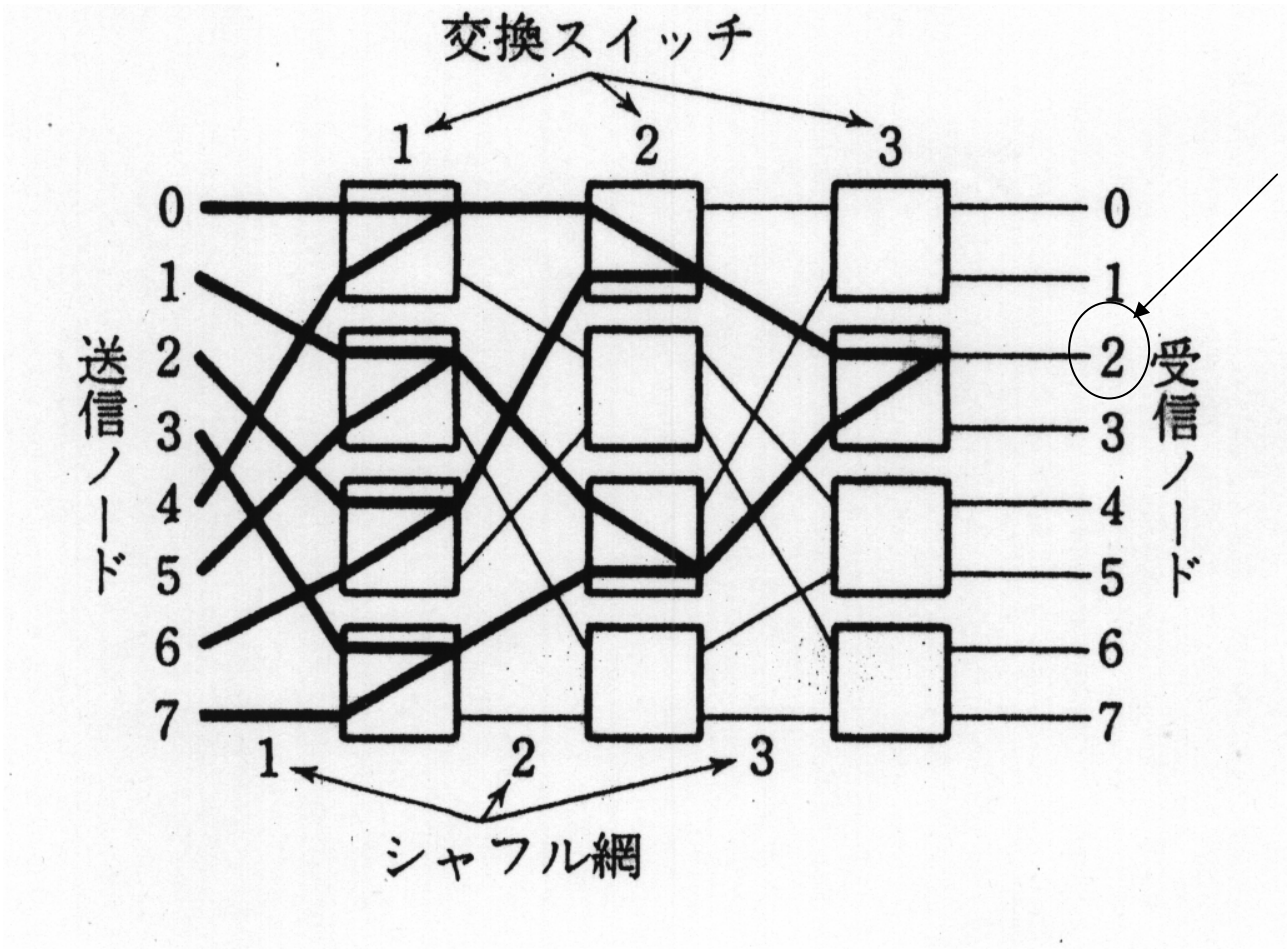
②可変トーラス網

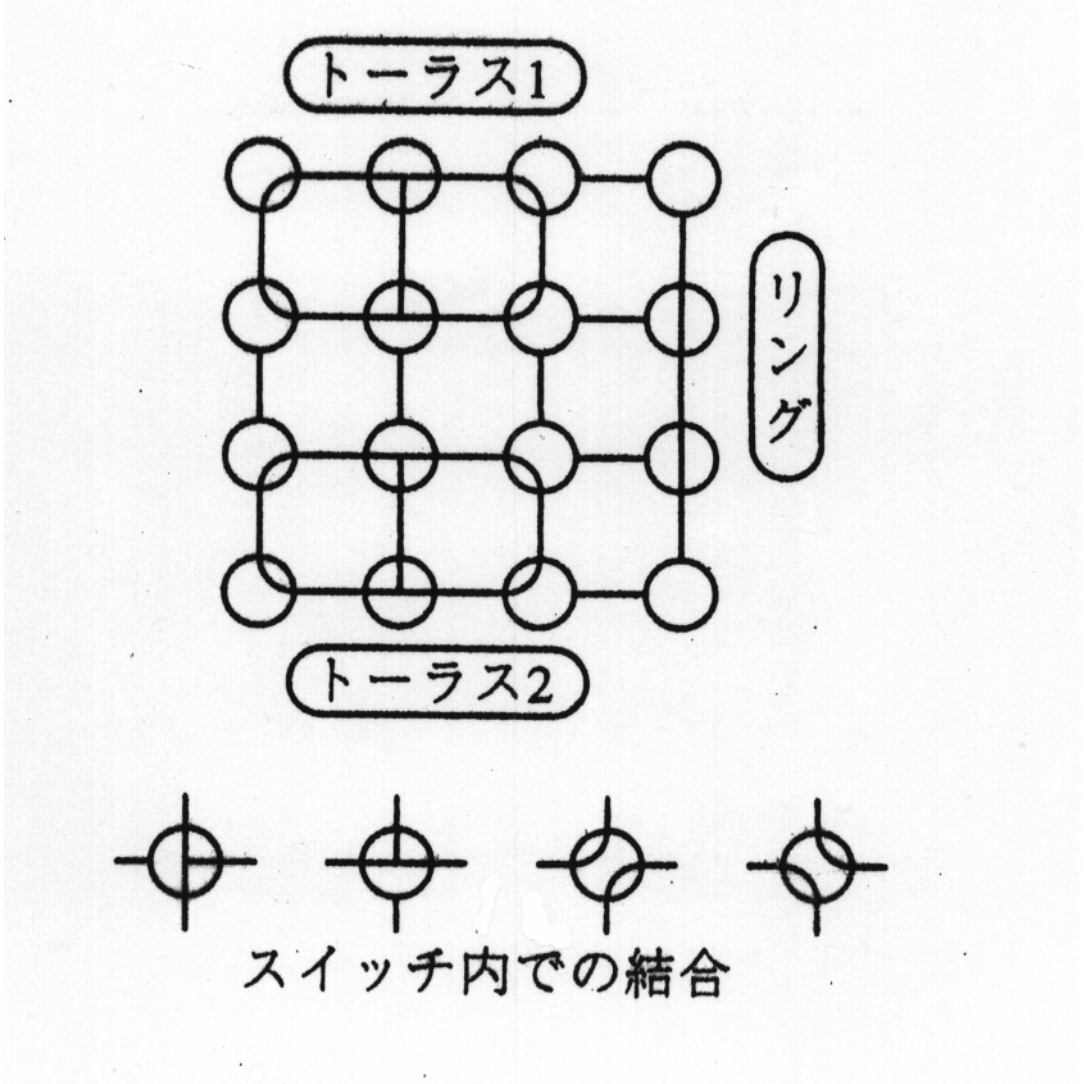
③多重クロスバ網

④高速網シミュレーション

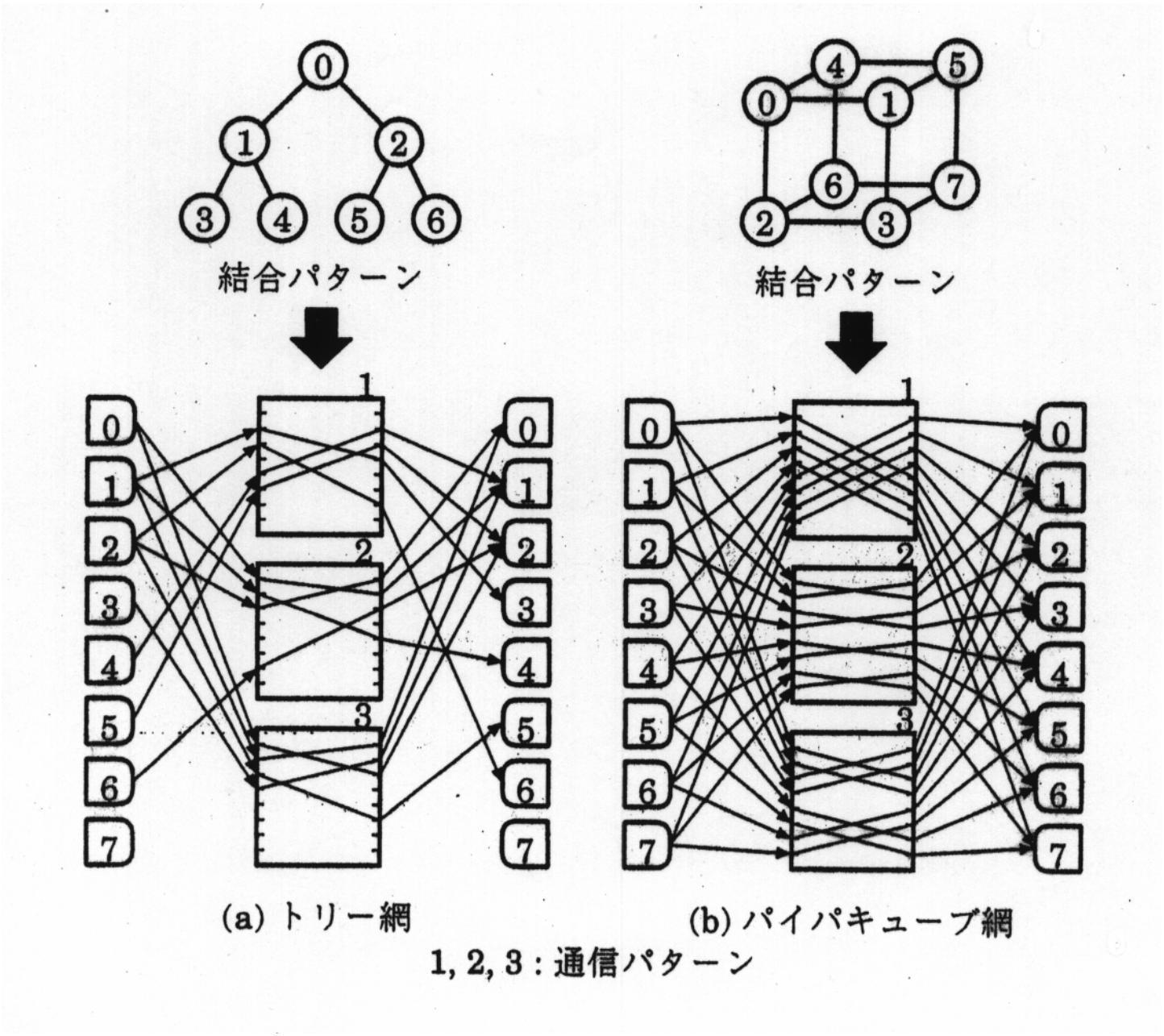
⑤動的リンク形成


送信プロセッサ側
負荷情報
Load$_k$0
Load$_k$1
データ 0
データ 1
Min(Load$_{k+1}$0,1)
比較器
Load$_{k+1}$0
Load$_{k+1}$1
負荷情報
データ
出力 0
データ
出力 1
受信プロセッサ側


交換スイッチ
1
2
3
0
1
2
3
4
5
6
7
送信ノード
受信ノード
制御変数格納
シャフル網



スイッチ内での結合


0
1
2
3
4
5
6
結合パターン
4
5
0
1
6
7
2
3
結合パターン
1
2
3
0
1
2
3
4
5
6
7
(a) トリー網
(b) パイパキューブ網
1, 2, 3 : 通信パターン