

シストリックアルゴリズムの 定式化とその結線構造

東北大学工学部 阿曾弘具 (Hiroতো Aso)

名古屋大学工学部 稲垣康善 (Yasuyoshi Inagaki)

1. まえがき

シストリックアルゴリズムは演算の並列性・規則性, 通信の局所性, 構造の一様性・拡張性に優れ, VLSI化に向いている。(1)-(4)

これまで, 具体的問題やある問題のクラスが与えられたとき, いかにシストリックアルゴリズムを構成するかが論じられてきた。(1)-(14) しかし, シストリックアルゴリズムで解きうる問題はどのようなものか, と問うことも必要であろう。これを考察の対象とする第一歩として, 本論文では, シストリックアルゴリズムを数学的対象として厳密に定式化する。

シストリックアルゴリズムの定式化はより一般的な設計法を記述するためにいくつか考えられている。文献(5)では有限順序機械をベースにして定式化されている。文献(6), (8)-(14)では設計法に重点をおいているため, シストリックアルゴリズムの定式化という点については直観に頼っているところがある。また, 1次元セル構造オートマトンとの関係を論じるための定式化もなされている。(7) 本論文の定式化は, それらに較べて数学的に厳密でより一般的なものとなっている。

本論文では, さらに, この定式化のもとで, 等価性を定義する。また, 結線構造がいくつかの型に分類できることを示し, 単純記憶型という型に等価変換できる条件を与える。

2. 定式化

2.1 準備

Z は整数の集合, $Z_p = \{0, 1, 2, \dots, p-1\}$, $Z^+ = \{0, 1, 2, \dots\}$ とする。

集合 D_1, D_2 に対して, D_1 から D_2 への (全域) 写像 $d: D_1 \rightarrow D_2$ の集合を $[D_1 \rightarrow D_2]$ と表し, 部分写像の集合を $[D_1 \rightarrow D_2]$ と表す。 D_1 が有限順序集合であるとき, $|D_1| = n$ (元の数) として, $[D_1 \rightarrow D_2]$ を n 次の直積空間 D_2^n と同一視する。すなわち, $D_1 = \{i_1, i_2, \dots, i_n\}$ であるとき, $d \in [D_1 \rightarrow D_2]$ は, $d = \langle d(i_1), d(i_2), \dots, d(i_n) \rangle$ とみなし, この右辺を $\langle d(i) \mid i \rangle$, $\langle d(i) \mid i \in D_1 \rangle$ とも記す。

シストリックアレイは普通1次元, または, 2次元空間に配置されるが, 本論では, 3次元空間 Z^3 ないし n 次元空間 Z^n , さらに, Z_p ($p > 1$, 1次元巡回型), $Z_p \times Z_q$ ($p, q > 1$, トーラス面), $Z_p \times Z$ (円柱面) などを想定して, 可換群 $G = \langle G; +, 0 \rangle$ に一般化して考える。可換群 G は Z -加群ともみなされる。すなわち, $s \in Z$, $\nu \in G$ に対して, $s\nu$ は s 個の ν の和として定まる G の元である。有限集合 $N \subseteq G$ について, N から生成される Z -加群を $\text{Span}(N)$ と記す。すなわち,

$N = \{v_1, v_2, \dots, v_n\}$ のとき, $\text{Span}(N) = \{\alpha \mid \alpha = \sum_{i=1}^n s_i v_i, s_i \in \mathbb{Z}\}$. $\text{Span}(N)$ は G の正規部分群をなす. 従って, G の $\text{Span}(N)$ による剰余群が存在し, それを $\text{Orth}(N)$ と記す. $\text{Orth}(N)$ は, $\alpha - \beta \in \text{Span}(N)$ のとき, $\alpha \equiv \beta$ と同値関係を定めるとその同値類の代表元からなる集合として, 表現できる. そうしたとき,

$$G = \text{Span}(N) + \text{Orth}(N)$$

である (+ は要素ごとの和を表す).

G とその有限集合 $N \subseteq G$ から, 有向グラフ $\langle G, E(N) \rangle$ を次のように定め, 結線グラフと呼ぶ.

$$E(N) = \{(\alpha, \alpha + v_i) \mid v_i \in N\}.$$

$\langle G, E(N) \rangle$ において $\text{Span}(N)$ を節点とする部分グラフは連結であり, $\text{Span}(N)$ による同値類が $\langle G, E(N) \rangle$ の連結成分を与える.

2.2 シストリックアルゴリズム

シストリックアルゴリズムは, 空間的広がりをもついくつかの種類のデータを対象として, 演算を実行する基本素子のある空間上に規則的に配置し, お互いの中で情報をやりとりして, データを処理していくものである. 従って, 基本セルの機能, セル間の結線などを定めるシストリックアレイと, 境界の入力セル群へどの入力データをいつ入力するか, 出力セル群からの各時点での出力のどれを実際出力データと認めるかを定めるタイミング関数とを定式化することで, シストリックアルゴリズムが定式化できる. まず, シストリックアレイを以下のように定式化する.

【定義1】シストリックアレイ SA は, セル機能 F , セル空間 G , セル間結線 W , 配置領域 R の4項組 $\langle F, G, W, R \rangle$ である. それぞれの定義と意味は次のようである.

(1) $F = \langle P_{in}, P_{out}, D, f \rangle$. P_{in} , P_{out} は有限順序集合で, 入力・出力ポート集合と呼ぶ. $|P_{in}| = n$, $|P_{out}| = m$ とおく. D は処理対象であるデータの集合で特別な元 \perp を含むものとする. \perp を静止データという. f は次の写像である.

$$f: P_{out} \rightarrow [[P_{in} \rightarrow D] \rightarrow D].$$

すなわち, $f = \langle f_1, f_2, \dots, f_a, \dots, f_m \rangle$ は $f_a: [P_{in} \rightarrow D] \rightarrow D$ なる写像の組である. 各 a について, $f_a(\perp, \dots, \perp) = \perp$ とする. $[P_{in} \rightarrow D] = D^n$ は入力ポート上のデータの集合を表し, f_a はその入力データに対して出力ポートの値を決める関数で, セル関数と呼ぶ. P_{in} の元を i, j で, P_{out} の元を a, b などで表す.

(2) G はある可換群 $G = \langle G; +, 0 \rangle$ である. G はセル機能を配置する空間の座標の集合であり, その元を α, β などで表し, セル座標あるいは単にセルと呼ぶ.

(3) W は, ポート間結線写像 $u: P_{in} \rightarrow P_{out}$ と 近傍(写像) $N: P_{in} \rightarrow G$ との2項組 $\langle u, N \rangle$ である. 各セル α の入力ポート i をセル $\alpha - N(i)$ の出力ポート $u(i)$ に接続することを意味する. $a \in P_{out}$ について, $u^{-1}(a)$ は出力ポート a からデータを受け取る入力ポートの集合を与える. N は順序集合 $\langle N(i) \mid i \in P_{in} \rangle$ も意味する.

(4) R は, G の空でない(有限)部分集合であり, アレイ領域とも呼ぶ. 実際にセル機能をおくセル座標の集合である.

以上, シストリックアルゴリズムのハードウェアともいべきシストリックアレイを定義した. セル空間 G は前節に述べたように普通に考えられているものを一般

化してある。実現しうるシストリックアレイにおける配置領域 R は有限集合で連結 (G と近傍 N から定まる結線グラフの上で連結) である。しかし、理論的考察のための便宜上無限集合または非連結である R を考えることがある。セルの機能を4項組で定めたが、 \perp は初期状態などまだ計算が始まっていないときのデータを表し、 $f_a(\perp, \dots, \perp) = \perp$ はすべての入力ポートに実際にデータがないときには出力ポートにもデータが現れないことを意味する。セル間結線について、ポート間結線写像 u を部分関数とすれば、他のセルの出力ポートに接続されない入力ポートが残り、それらをブロードキャスト入力ポートともみなせる。しかし、本論では全域関数であるものだけを考える。また、 u が全射でなければ、他のセルの入力ポートから参照されない出力ポートがあることを意味する。アレイの動作を考察する立場からは、それらは取り除いてしまってもよいだろう。以下では、 u は全射であると仮定する。本定式化では、通信の局所性、構造の一様性を重視して、データのブロードキャストは行わず、一部に特別な機能をもつセルを使うことはしないものとした。

アレイへの入出力は、各セルの入出力ポートでその接続先がアレイの外に出てしまうものをアレイに対する入出力ポートとみなすことによって実現できるであろう。このアレイ入力ポートに入力データを次々に与えることで、シストリックアレイは動き出す。この動作を次のように定める。 $T = \{0, 1, 2, \dots\}$ を時刻の集合とする。

【定義2】(1) SAにおいて、アレイ入力・出力ポート集合 PA_{in} , PA_{out} を次のように定める。

$$PA_{in} = \{(i, \alpha) \mid \alpha \in R \text{ かつ } \alpha - N(i) \notin R\}$$

$$PA_{out} = \{(a, \alpha) \mid \alpha \in R \text{ かつ } \exists i \in u^{-1}(a), \alpha + N(i) \notin R\}.$$

$$(PA_{in} \subseteq P_{in} \times R, PA_{out} \subseteq P_{out} \times R).$$

(2) 各セルの各ポート上にあるデータの組の全体を様相という。すなわち、 $[P_{in} \times R \rightarrow D]$ をセル入力様相集合、 $[P_{out} \times R \rightarrow D]$ をセル出力様相集合という。また、アレイ入出力ポート上に制限した様相、すなわち、 $[PA_{in} \rightarrow D]$, $[PA_{out} \rightarrow D]$ をアレイ入力様相集合、アレイ出力様相集合という。

(3) SAへの入力データ流とは、 $d_{in}: T \rightarrow [PA_{in} \rightarrow D]$ であり、出力データ流は、 $d_{out}: T \rightarrow [PA_{out} \rightarrow D]$ である。 $d_{in}(t)$, $d_{out}(t)$ を d_{in}^t , d_{out}^t と記す。

(4) SAの動作とは、入力データ流 $\{d_{in}^t\}$ が与えられたとき、次のように定まるセル入力様相の列 $y^0, y^1, \dots, y^t, \dots$, セル出力様相の列 $x^0, x^1, \dots, x^t, \dots$, 出力データ流 $d_{out}^0, d_{out}^1, \dots, d_{out}^t, \dots$ である。

$$y^0(i, \alpha) = \perp \quad [\forall (i, \alpha) \in P_{in} \times R]$$

$$y^t(i, \alpha) = \begin{cases} d_{in}^{t-1}(i, \alpha) & [(i, \alpha) \in PA_{in}, t > 0] \\ x^{t-1}(u(i), \alpha - N(i)) & [(i, \alpha) \notin PA_{in}, t > 0] \end{cases}$$

$$x^0(a, \alpha) = \perp \quad [\forall (a, \alpha) \in P_{out} \times R]$$

$$x^t(a, \alpha) = f_a(\langle y^t(i, \alpha) \mid i \in P_{in} \rangle) \quad [(a, \alpha) \in P_{out} \times R, t > 0]$$

$$d_{out}^t(a, \alpha) = x^t(a, \alpha) \quad [(a, \alpha) \in PA_{out}]$$

(5) SAの機能は、上記の動作で定まる入力データ流 d_{in} から出力データ流 d_{out} への対応 SA_{io} である。

$$d_{out} = SA_{io}(d_{in})$$

シストリックアレイの拡張性を表現するため、次の定義をする。

【定義3】(1) SAにおいて、連結用入力ポート集合とは次の PA_{cin} である。

$$\begin{aligned} PA_{in} &= \{(a, \beta) \mid \exists i \in u^{-1}(a), (i, \beta + N(i)) \in PA_{in}\} \\ &= \{(u(i), \alpha - N(i)) \mid (i, \alpha) \in PA_{in}\}. \end{aligned}$$

(2) $d_{in} : T \rightarrow [PA_{in} \rightarrow D]$ を 連結用入力データ流 という.

PA_{in} はアレイ領域 R の境界のすぐ外にあるセルの出力ポートに対応している. 2つの SA: A_1, A_2 について, A_1 の PA_{out} と A_2 の PA_{in} との共通部分が空でないならば, その共通部分のポートを通じて2つを接続できる.

連結用入力データ流から次のようにして入力データ流が決まる.

$$d_{in}(t, i, \alpha) = d_{in}(t, u(i), \alpha - N(i)).$$

このデータ流を用いて, SAの機能を次のようにも定めることができる.

$$d_{out} = SA_{00}(d_{in})$$

シストリックアルゴリズムが処理対象とするデータは空間的広がりをもつものであった. この空間をデータ空間と呼び, G_D と記す. G_D は1次元, または, 2次元格子空間とするのが普通であるが, ここではセル空間を可換群とした理由と同じ理由で G_D を可換群とだけ規定しておく. (処理対象データ毎に空間を定義した方が一般的であるが, 煩雑になるので本論文では G_D 1つに限ることにした). 処理対象データは $[G_D \rightarrow D]$ の元として定式化できる. G_D 上の各点の値をSAのどのアレイ入力ポートにいつ入力するか, どの出力ポートのいつの時点での値を G_D 上のどの点の値とみなすか, を定めることがタイミング関数を与えることである. すなわち, シストリックアルゴリズムは次のように定式化できる.

【定義4】(1) シストリックアルゴリズムとは3項組 $\langle A, G_D, \text{Timing} \rangle$ である. ただし, A はシストリックアレイ, G_D は可換群でデータ空間, Timing は $\{I_1, I_2, \dots, I_p, \dots, O_1, O_2, \dots, O_q, \dots\}$ なる有限集合である. ここで, $I_p \in [G_D \rightarrow PA_{in} \times T]$, $O_q \in [G_D \rightarrow PA_{out} \times T]$ であり, $\{I_p(G_D)\}$, $\{O_q(G_D)\}$ はそれぞれ $PA_{in} \times T$, $PA_{out} \times T$ の排反な部分集合の族になっているものとする. I_p は p 番目の入力データのタイミングを, O_q は q 番目の出力データのタイミングを表し, Timing を タイミング関数族 と呼ぶ.

(2) 入力データ $input_p \in [G_D \rightarrow D]$ が与えられたとき,

$$d_{in}(t)(i, \alpha) = \begin{cases} input_p(\mu) & [I_p(\mu) = \langle i, \alpha, t \rangle] \\ \perp & [\text{その他}] \end{cases}$$

と入力データ流を定め, SAの出力データ流 $d_{out} = SA_{10}(d_{in})$ から出力データ $output_q \in [G_D \rightarrow D]$ を次のように定める.

$$output_q(\mu) = d_{out}(t)(a, \alpha) \quad [O_q(\mu) = \langle a, \alpha, t \rangle].$$

以上でシストリックアルゴリズムの定式化が完了した. 具体例を1つ示す.

例1. 2つの $n \times n$ 行列 a, b の積を $n \times n$ 行列 c に求めるシストリックアルゴリズムの1つ⁽¹²⁾を示そう.

シストリックアレイは次の通りである. (u^{-1} , IN については3.2参照.)

$$P_{in} = \{1, 2, 3\}, \quad P_{out} = \{a, b, c\}$$

$$D = \text{数の集合} \cup \{\perp\}$$

$$f = \langle f_a, f_b, f_c \rangle$$

$$f_a(d_1, d_2, d_3) = d_1$$

$$f_b(d_1, d_2, d_3) = d_2$$

$$f_c(d_1, d_2, d_3) = d_3 + d_1 \times d_2$$

$$G = Z \times Z,$$

P_{in}	u	N	P_{out}	u^{-1}	IN
1	a	$(-1, 0)$	a	{ 1 }	{ 1 }
2	b	$(0, -1)$	b	{ 2 }	{ 2 }
3	c	$(1, 1)$	c	{ 3 }	{ 1, 2, 3 }

$$R = \{1, 2, \dots, size\} \times \{1, 2, \dots, size\}$$

(size は奇数とし, $size = 2z-1$, $z \geq n$ とする).

タイミング関数族 $Timing = \{I_1, I_2, O_3\}$ は次のように定める. $G_D = Z \times Z$ とし, 各タイミング関数の定義域は $\{1, 2, \dots, n\} \times \{1, 2, \dots, n\}$ である. G_D の元は $(i, k), (k, j), (i, j)$ で示す.

$$I_1(i, k) = \langle 1, (size, z-i+k), i+2k-2 \rangle$$

$$I_2(k, j) = \langle 2, (z-j+k, size), j+2k-2 \rangle$$

$$O_3(i, j) = \begin{cases} \langle c, (size+i-j, size), 2i+j-3+size \rangle & [i \leq j] \\ \langle c, (size, size-i+j), i+2j-3+size \rangle & [i > j] \end{cases}$$

データ $a(i, k), b(k, j)$ を I_1, I_2 で入力し, O_3 でデータ $c(i, j)$ が得られる.

この例は, $|i-k| < z, |k-j| < z$ なる添字 $(i, k), (k, j)$ にデータをもつ帯行列の積を求めるシストリックアルゴリズム⁽¹⁾ともみなせる.

3. シストリックアレイの構造

3.1 シストリックアレイの機能

アレイ領域 R の境界ではアレイ入出力ポートを通じてデータが入出力するが, R の境界を除く内部のセルのポート a 上のセル出力様相は,

$$x^t(a, \alpha) = f_a(\langle x^{t-1}(u(i), \alpha - N(i)) \mid i \in P_{in} \rangle)$$

と定まる. すなわち, 各時点でのセル出力様相から次の時点でのセル出力様相が定まる. これはシストリックアレイの処理対象情報がセル出力様相であるとしてよいことを意味するだろう. 従って, シストリックアレイの本質的な機能はセル出力様相の変換であると言えるだろう. このことは次のように定式化できる.

【定義5】(1) $R = G$ とした場合のセル出力様相を単に様相と呼び, その集合を C と記す. すなわち,

$$C = \{x \mid x : P_{out} \times G \rightarrow D\}.$$

(2) SA に対して, C 上の変換 τ を次のように定め, SA の並列写像という. 任意の $x \in C$ について,

$$\tau(x)(a, \alpha) = f_a(\langle x(u(i), \alpha - N(i)) \mid i \in P_{in} \rangle)$$

従って, $R = G$ の場合, $x^t = \tau(x^{t-1})$ であり, τ が SA の本質的機能を与えるとみなすことができる. また, SA がセル出力の組を状態とするセル構造オートマトンであることを示している. しかし, 状態がセル出力の組という構造をもったものとなり, その状態遷移関数も内部構造をもつので, セル構造オートマトンの一般論では議論できないことが議論できる.

さて, 写像 $\gamma : P_{out} \rightarrow G$ に対して, 写像 $s^\gamma : C \rightarrow C$ を次のように定め, (非一様)シフトという.

$$s^\gamma(x)(a, \alpha) = x(a, \alpha - \gamma(a))$$

すなわち, 各出力ポート a における様相 $x(a, *)$ を $\gamma(a)$ 方向に移動する写像であ

る。 γ が定値写像のとき 一様シフト という。シフト s^{γ} は全単射であり、その逆写像は s^{γ} で与えられる。従って、 $s^{\gamma}(x) = x'$ となる x と x' とは本質的には同じものであるとみなせる。これを基礎にした次の定義は妥当であろう。

【定義6】 2つのSAに対して、その並列写像を τ, τ' とする。ある γ が存在して、 $\tau'(s^{\gamma}(x)) = s^{\gamma}(\tau(x))$ のとき、2つのSAは 等価 であるという。 s^{γ} が一様シフトであるとき、強等価 であるという。

すなわち、一方のSAは x を $\tau(x)$ に変換し、他方のSAは x' を $\tau'(x')$ に変換する。このとき、 x' が x をシフトしたものであるならば $\tau'(x')$ が $\tau(x)$ をシフトしたものととなるとき、かつそのときに限り、2つのSAは等価である。

【補題1】 任意のSAの並列写像 τ と任意の γ に対して、 $\tau'(s^{\gamma}(x)) = s^{\gamma}(\tau(x))$ を満たす C 上の写像 τ' をその並列写像とするSAが存在する。

【証明】 $\tau'(x) = s^{\gamma}(\tau(s^{\gamma}(x)))$ でなければならない。

$$\begin{aligned} \tau'(x)(a, \alpha) &= \tau(s^{\gamma}(x))(a, \alpha - \gamma(a)) \\ &= f_a(\langle s^{\gamma}(x)(u(i), \alpha - \gamma(a) - N(i)) \mid i \rangle) \\ &= f_a(\langle x(u(i), \alpha - \gamma(a) - N(i) + \gamma(u(i))) \mid i \rangle) \end{aligned}$$

そこで、次のように定めて、目的のSAが得られる。

$$P'_{in} = P_{in} \times P_{out}.$$

各 $(i, b) \in P'_{in}$, $d \in [P'_{in} \rightarrow D]$ に対して、

$$\begin{aligned} u'(i, b) &= u(i), \quad N'(i, b) = N(i) - \gamma(u(i)) + \gamma(b), \\ f'_a(\langle d(i, b) \mid (i, b) \rangle) &= f_a(\langle d(i, a) \mid i \rangle) \quad \square \end{aligned}$$

【命題1】 任意のSAに対して、そのセル間結線 W が次の条件をみたす強等価なSAが存在する。

条件：任意の $i, j \in P_{in}$ について、 $i \neq j$ ならば、 $u(i) = u(j)$ かつ $N(i) = N(j)$ となることはない。

【略証】 $i, j \in P_{in}$ に対して、同値関係 \equiv を、

$$i \equiv j \iff u(i) = u(j) \text{ かつ } N(i) = N(j)$$

と導入する。この同値関係により P_{in} を同値類分割し、その代表元を集めたものを新しい入力ポート集合とすることで、目的のSAが得られる。 \square

これは、接続先が同じ入力ポートは同一視してよいことを示し、異なる入力ポートの接続先は異なるとしても一般性を失わないことを示している。

3.2 情報の流れと結線構造

シストリックアレイ上の様相がいかなる情報をもとに変換されていくかを考察する。

セル写像 f_a は、一般にはすべての入力には必ずしも依存しないであろう。そこで、各 $a \in P_{out}$ について、

$$IN(a) = \{ i \mid f_a \text{ は } i \text{ に依存する} \}$$

と定める。 f_a が i に依存するとは、

$$f_a(d_1, d_2, \dots, d_i, \dots, d_n) \neq f_a(d_1, d_2, \dots, d_i', \dots, d_n)$$

かつ $d_i \neq d_i'$ となる $d_1, d_2, \dots, d_i, \dots, d_n, d_i'$ が存在することである。

各 b, a について、 $P(b, a) = u^{-1}(b) \cap IN(a)$ とおく。 $i \in P(b, a)$ であることは、セル $\alpha - N(i)$ の出力ポート b の情報がセル α の入力ポート i を通じて f_a による変換を受けその出力ポート a へ流れることを意味する。この情報の流れの

道は次の有向グラフで表現できる。

【定義7】 SAに関するデータフローネットワーク DFN とは、有向グラフ $\langle P_{out} \times R \cup P_{a_{in}}, E(f, u, N) \rangle$ である。ここで、辺集合 $E(f, u, N)$ は、次で定める。

$$E(f, u, N) = \{ \langle (b, \alpha - N(i)), (a, \alpha) \rangle \mid i \in P(b, a) \}$$

DFN の $P_{a_{in}}$ の元を始点とし、 $P_{a_{out}} \subseteq P_{out} \times R$ の元を終点とするパスにそって、情報が流れると考えられる。各パスを構成している辺は入力ポート名でラベルづけされているともみなす。

次は定義7より明らかである。

【補題2】 DFNにおいて、 (b, β) から (a, α) へのパスが存在するならば、

$$\alpha - \beta \in \text{Span}(N)$$

である。ただし、近傍 N を集合と同一視している。

従って、 $\alpha - \beta \notin \text{Span}(N)$ ならばセル α と β とはつながっていない。セル α に関わる情報の流れを問題にすると、 α が属する連結成分内のみで議論を進めればよい。従って、 $G = \text{Span}(N)$ であると仮定して一般性を失わない。

DFN は、サイクルをもつ場合と、もたない、すなわち、非巡回的な場合とがある。後者の場合、 (b, β) から (a, α) へのパスにおいてその途中に同じ節点が現れないから、並列に並んだ情報が一方向に流れながら処理されるといえる。一方、前者の場合、あるサイクルを $(a_0, \alpha_0), (a_1, \alpha_1), \dots, (a_{s-1}, \alpha_{s-1}), (a_0, \alpha_0)$ としたとき、 $\alpha_0 = \alpha_1 = \dots = \alpha_{s-1}$ が成立するならば、セル α_0 内の出力ポート間を情報が流れるだけで、セル α_0 が記憶をもって順序回路的に動作しているとみなせる。そうでない場合は、セル α_0 の出力ポート a_0 を出た情報は他のセルを通過して処理を受けた後、再び α_0 の a_0 への入力となることを意味し、サイクル内のセル全体で動的記憶をもつとみなせ、それらが協調的に働いて処理していることを意味する。

【定義8】 SAに関するDFNにおいて、サイクル $(a_0, \alpha_0), (a_1, \alpha_1), \dots, (a_{s-1}, \alpha_{s-1}), (a_0, \alpha_0)$ が存在する場合にはつねに $\alpha_0 = \alpha_1 = \dots = \alpha_{s-1}$ が成立するとき、単純記憶型SAという。また、DFNが非巡回的である場合には、パイプライン型SAという。

例1のSAはパイプライン型SAである。

このDFNにおけるパス、従って、SAにおける情報の流れを分析するために次のグラフを定義する。

【定義9】 (1) シストリックアレイに対するデータ依存グラフDDGとは、節点集合 V_d を $P_{out} \cup P_{in}$ 、辺集合 $E_d \subseteq P_{out} \times P_{in} \cup P_{in} \times P_{out}$ を次のように定め、各辺 (b, i) には $N(i)$ というラベルを付けたラベル付2部グラフである。(辺 (i, a) にはラベルを付けない)。

$$E_d = \{ (b, i) \mid b = u(i), i \in P_{in} \} \cup \{ (i, a) \mid i \in \text{IN}(a), a \in P_{out} \}$$

(2) $a_0, a_1, \dots, a_{s-1}, a_s \in P_{out}$, $i_0, i_1, \dots, i_{s-1} \in P_{in}$ に対して、 $(a_t, i_t), (i_t, a_{t+1}) \in E_d$ のとき、 $c = \langle a_0, i_0, a_1, i_1, a_2, \dots, a_{s-1}, i_{s-1}, a_s \rangle$ をDDGのパスといい、 $a_0 = a_s$ のときサイクルという。(一般には P_{in} の元を始点または終点とするものもパスというが本論文では上のものに限る)。

(3) このDDGのパス c に対して、

$$\delta(c) = \sum_{t=0}^{s-1} N(i_t)$$

とおき, c に沿う相対位置という。

(4) DDG において, $\delta(c)=0$ であるサイクル c で, その部分サイクル c' として $\delta(c')=0$ となるものを含まないとき, ゼロサイクルという。

このDDG とDFN との関係は次の補題で示される。

【補題3】 DFN において, (b, β) から (a, α) へのパスが存在する必要十分条件は, DDG において b から a へのパス c が存在して, $\delta(c)=\alpha-\beta$ となることである。

【略証】定義7, 9より容易に示される。 □

補題3より次の2つが成立する。

【補題4】 DFN においてサイクルが存在する必要十分条件は, DDG においてゼロサイクルが存在することである。

【補題5】 DFN が連結であるならば, DDG も連結である。

従って, DDG は連結であると仮定しても一般性を失わない。

DDG の節点集合を強連結性で同値類に分け, 各同値類を層(layer)という。各層を1つの節点とみなすグラフは非巡回的であり, 出力ポート間の一方向の情報の流れ, すなわち, 各層で処理される情報の依存性を表現している。このグラフを層間データ依存グラフと呼ぶ。層間データ依存グラフから, 層を次のように分類できる。はいる辺がない層, 出る辺がない層, 共にある層の3つである。それぞれ, 入力層(アレイ入力が必要な層), 出力層(アレイ出力が必要な層), 中間処理層と名付けることができる。それぞれの層にはアレイ入出力があるかも知れない。

層のみ取り出したグラフを一般層内データ流グラフと呼ぶ。一般層内データ流グラフにおける各節点 a に対して, a を含むサイクルの集合を $CF(a)$ とおき, a -情報流候補集合とよぶ。この $CF(a)$ によりシストリックアレイにおける出力ポート a の役割を次のように分類できる。 $CF(a) = \phi$ ならば, a は層間ラッチであり, 情報を1単位時間遅延させて次の層に送る役割を果たす。 $CF(a) \neq \phi$ のときは, a を層内データ流といい, さらに分ける。 $CF(a)$ がゼロサイクルを含まないならば, a を出た情報は何時か後に他のセルのポート a に現れるとみなせ, 層内の情報は一方向の流れ(層内流)を形成していると言える。 $CF(a)$ がゼロサイクルを含むならば, DFN のあるサイクルの中に a があり, ポート a は動的記憶を担っている。

この動的記憶の比較的扱い易いものとして単純記憶型SAを定義した。次に, その単純記憶型SAに等価変換できる条件を与える。まず, グラフのサイクルのある集合について次の概念を定める。

【定義10】 (1) サイクルのある集合 Cyc に対して, 節点集合を Cyc とし, 2つの Cyc の元が1点以上を共有するときその2元間に辺を定め, 1つのサイクルで交叉しているときセルフループを定めたグラフを, Cyc の随伴グラフという。

(2) サイクルのある集合 Cyc がカクタス(cactus)であるとは, その任意の2つの元のサイクルで共有される点が1点以下であり, その随伴グラフが木の集合であることである。

次は, 単純記憶型の定義および補題3から明らかである。

【補題6】 SAが単純記憶型である必要十分条件は, そのDDGのすべてのゼロサイクルの上の辺のラベルが0であることである。

【定理1】 シストリックアレイに対して, そのDDGのすべてのゼロサイクルから

なる集合がカクタスであるならば、それと等価な単純記憶型SAが存在する。

〔証明〕 DDG の1つのゼロサイクル $c = \langle a_0, i_0, a_1, i_1, a_2, \dots, a_{s-1}, i_{s-1}, a_s \rangle$ に対して、 $\gamma(a_i) = \sum_{p=i}^{s-1} N(i_p)$ とし、他の a については $\gamma(a) = 0$ とする。補題1より、この γ に対してSAが存在する。そのSAのDDGの対応するゼロサイクル上の辺のラベルは0となる。辺のラベルが0ではないゼロサイクル c' があれば、上と同様に新しい γ を定める。ただし、 c' と点を共有する辺のラベルが0であるゼロサイクル上の各点 b の γ の値はその共有点の γ の値と等しくする。カクタスであるという条件よりこれは可能である。この γ に基づいて補題1の証明と同様の方法で等価な新しいSAを構成できる。辺のラベルが0ではないゼロサイクルがある限り、これを繰り返す。補題6より、最終的に得られたSAが単純記憶型SAである。□

【系1】 SAのDDGのゼロサイクルの部分集合でカクタスとなるものに対して、そのゼロサイクル上の辺のラベルを0とする等価なSAを構成できる。

4. むすび

本論文では、シストリックアルゴリズムを定式化し、情報の流れを記述解析する方法を与えた。また、シストリックレイの型として単純記憶型を導入し、それに等価変換できるための条件を与えた。

シストリックアルゴリズムを適用できる問題は、対象データがいくつかの類に分かれ、それぞれが空間的広がりを持ち、データ処理が空間内の局所的処理の積み重ねでなされものである必要がある。この空間的広がりをもつデータをシストリックアルゴリズムの入出力データ流にいかに変換するかは、設計上の重要な点であり、情報の流れ方及びデータの出会いの解析によりその変換を与えることができるが、機会をあらためて報告する⁽¹⁵⁾。Moldovan はループプログラムで記述された問題の対象データの依存関係からシストリックアルゴリズムを合成する手法を与えており⁽⁶⁾⁽¹³⁾、本論文の定式化によりその方法をより組織的な形で実現できるものと思われるが、その詳しい議論は今後に残された課題である。

謝辞 本研究を進めるうえで御討論頂いた豊橋技術科学大学本多波雄学長、東北大学木村正行教授、東海情報論談話会の皆様、並びに名古屋大学稲垣研究室の皆様へ感謝する。また、本研究は一部、文部省科学研究費補助金(一般(C)課題番号60550263)の援助を受けた。ここに記して感謝する。

文 献

- (1) H. T. Kung, C. E. Leiserson; "Systolic array (for VLSI)", Proc. Symp. Sparse Matrix Computations and Their Applications, pp. 256-282. (1978)
[pp. 271-292 of "Introduction to VLSI systems" by C. A. Mead, L. A. Conway, Addison-Wesley (1980)]
- (2) U. Weiser, A. Davis: "A wavefront notation tool for VLSI array design", VLSI Systems and Computations, Computer Science Press, pp. 226-234 (1981)
- (3) S. Y. Kung, K. S. Arun, R. G. Gal-Ezer, D. V. B. Rao: "Wavefront array processor: language, architecture, and applications", IEEE Trans. Comp., C-31, 11, pp. 1054-1066 (1982)

- (4) H. T. Kung; "Why systolic architecture?", IEEE Computer, 15, 1, pp. 37-46. (1982)
- (5) C. E. Leiserson; "Area-efficient VLSI computation", MIT Press. (1983)
- (6) D. I. Moldvan; "On the design of algorithms for VLSI systolic arrays", Proc. IEEE, 71, 1, pp. 113-120. (1983)
- (7) 梅尾博司; "シストリックアレイによる線形時間限定セルラ・オートマトンの模倣", 電子通信学会論文誌D, J66-D, 6, pp. 715-721. (1983)
- (8) 若林, 菊野, 吉田; "漸化式を用いたハードウェアアルゴリズムの設計について", 電子通信学会論文誌D, J67-D, 8, pp. 861-868. (1984)
- (9) P. Quinton; "Automatic synthesis of systolic arrays from uniform recurrent equations", Proc. 11th Ann. Int. Symp. Computer Architecture, pp. 208-214. (1984)
- (10) M. S. Lam, J. Mostow; "A transformational model of VLSI systolic design", Computer, 18, 2, pp. 42-52 (1985)
- (11) 西田, 堀池, 坂口; "シストリックに結合されたVLSIプロセッサアレイへのタスク割り振りの一手法", 電子通信学会技術報告 CAS85-174. (1985)
- (12) 三浦, 阿曾, 稲垣; "多重ループプログラムを処理するシストリックアルゴリズムの構成法", 電子情報通信学会論文誌D (1987. 3. 以降掲載予定)
["シストリックアルゴリズム構成法の基礎", 電子通信学会技術報告 AL85-50. (1985. 12)]
- (13) D. I. Moldvan, J. A. B. Fortes; "Partitioning and mapping algorithms into fixed size systolic arrays", IEEE Trans. Comp., C-35, 1, pp. 1-12. (1986)
- (14) O. H. Ibarra, S. M. Kim, M. A. Palis; "Designing systolic algorithms using sequential machine", IEEE Trans. Comp., C-35, 6, pp. 531-542. (1986)
- (15) 阿曾, 稲垣; "シストリックアルゴリズムの定式化と情報の流れ", 電子情報通信学会論文誌D (1987. 6. 以降掲載予定)