

## 正規集合を受理するパターン・マッチング・マシン

Pattern Matching Machine for Recognizing Regular Set

広島大学 工学部 若林 真一 (Shin'ichi WAKABAYASHI)

藤井 隆志 (Takashi FUJII)

菊野 亨 (Tohru KIKUNO)

### 1. まえがき

パターン・マッチングはパターンと呼ばれるあらかじめ設定された系列に対し、与えられた文字系列がパターンに一致するか否かを決定する操作であり、データベースにおけるキーワード検索などにおいて非常に重要である。一方、VLSI 技術の進歩に伴い、従来ソフトウェアで実現されていたアルゴリズムを専用プロセッサとして VLSI のチップ上に直接実現しようとするハードウェア・アルゴリズムあるいは VLSI アルゴリズムが注目されている<sup>(5)</sup>。

ここではパターンとして正規表現を仮定したパターン・マッチングに対するハードウェア・アルゴリズムについて議論する。既に Chu<sup>(1)</sup>, Floyd<sup>(2)</sup>, Foster<sup>(3)</sup>, Mukhopadhyay<sup>(6)</sup> のものが知られているが、これらにおいてはパターンを実行時に設定することができない。それに比べ本稿で提案するア

ルゴリズムはパターンを実行時に設定できる（これをプログラマブルであるという），更に，規則正しい構造のためVLSI化が容易である，という特長を持つ。提案するハードウェア・アルゴリズムをプログラマブル・パターン・マッチング・マシンと呼びPPMと略記する。

## 2. 準備

正規表現 $P$ ， $P$ で表される正規集合 $L(P)$ などは通常の意味，例えば文献(4)，に従う。アルファベットを $\Sigma$ で表し，正規表現には3つの演算子 $\cdot$ （接続）， $+$ （和集合）， $*$ （Kleene 閉包）を許すものとする。記述を簡単にするためワイルド文字 $?$ （ $\notin \Sigma$ ）を導入し， $? = a_1 + a_2 + \dots + a_{|\Sigma|}$ （ $a_i \in \Sigma$ ）とする。又，空系列を $\epsilon$ （ $\notin \Sigma$ ）で表す。任意の正規表現は解析木（parse tree）で表すことができる。

$\Sigma = \{a, b, c, d\}$  とし，正規表現を $P = a^* \cdot (b + \epsilon) \cdot (c + d \cdot ?)$  とする。このとき $L(P) = \{c, ac, aac, abc, \dots\}$  となり， $P$ の解析木は図1に示す通りである。

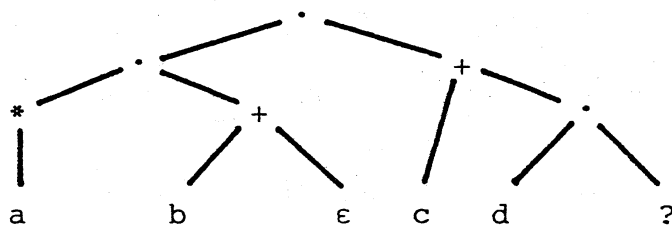


図1 解析木

ここではnMOS技術を前提とし、ハードウェア・アルゴリズムは同期式順序回路として実現されるものとする。そのためノンオーバーラッピング2相クロック $t_0, t_1$ を仮定し、クロック周期を $T$ とする。

### 3. 基本回路PM

ここではPPMの基本となる回路(PMと呼ぶ)の回路構成及びその基本動作について述べる。

#### 3.1 回路構成

パターンとして与えられた正規表現 $P$ に対し、PMは $P$ の解析木に基づいて構成される。例えば図1の解析木に対するPMは図2のようになる。すなわちPMは比較セル( $\square$ )と演算子セル( $\odot, \oplus, \otimes$ )で構成され、セル間は解析木に対応して接続される。比較セルはバスに接続され、入力系列はバスを介してPMに入力される。

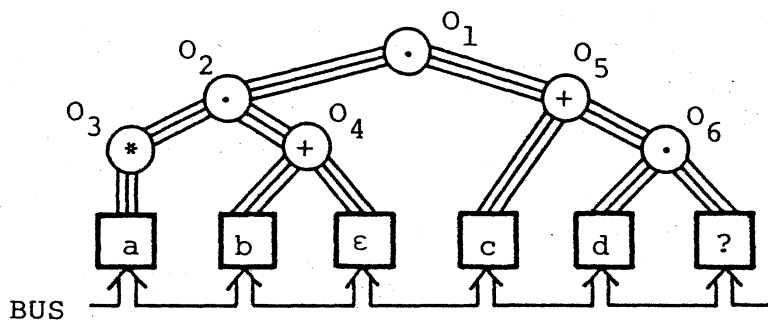


図2 PM

図3に比較セルと $\oplus$ セルの構成を示す(他の演算子セルについては文献(7)を参照されたい)。比較セルはレジスタR, 比較器CMP, ラッチD, 及びいくつかのゲートから構成される。レジスタRは正規表現に含まれる文字(?,  $\epsilon$ を含む)をラッチする。Rの端子 $\epsilon$ , ?はRが $\epsilon$ , 又は?をラッチしているとき1となる。CMPはR内のデータとバス上のデータを比較し, 一致したとき端子Cを1とする。次に $\oplus$ セルの演算子セルは和集合の演算を行う。

各セルの端子E,  $R_1$ ,  $R_2$ に接続されている信号線をそれぞれe,  $r_1$ ,  $r_2$ で表す。

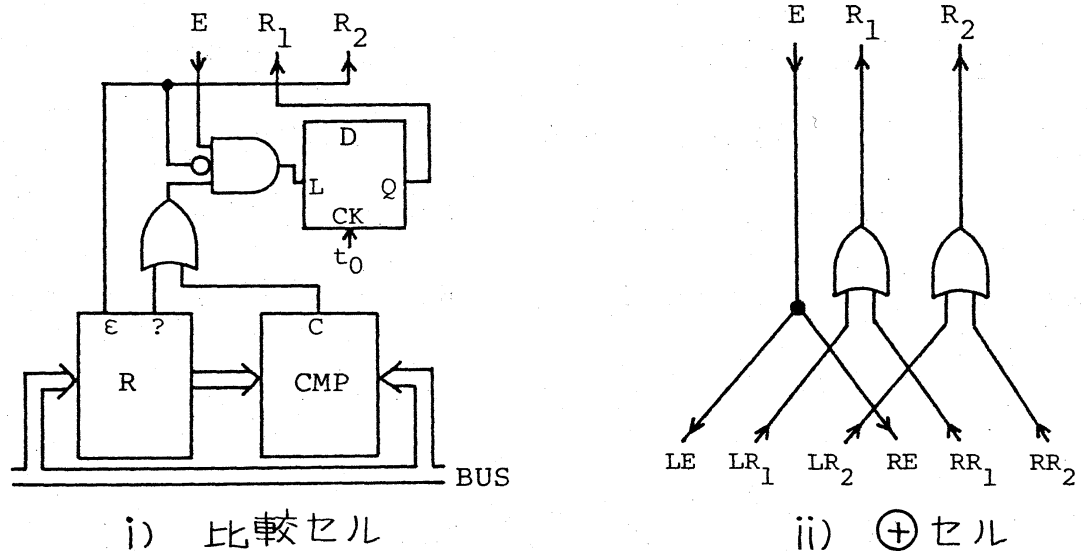


図3 各セルの回路構成

### 3.2 動作

各比較セルのレジスタRには既にパターンである正規表現

Pの文字がラッチされているものとする。入力系列 $S = a_1 a_2 \dots a_m$ の各文字はクロック $t_0$ に同期して、順次、バス上に入力される。 $a_1$ を入力するとき、解析木の根に対応する演算子セル(図2の $O_1$ )の端子Eを1とする。 $S$ とPが一致すれば、 $S$ の最後の文字 $a_m$ を入力したクロック周期において $O_1$ の端子 $R_1$ が1となる。すなわちPMはオンラインで動作する。

図2のPMに $S = abc$ を入力した場合のPMの動作の様子を図4に示す。図4の $\rightarrow$ はセル間の $e, r_1$ 上の信号の流れを示す。この信号はどの比較セルが実際に比較を行うかを指示する。 $\circ$ はそのクロック周期に比較を行うセル、 $\odot$ は比較の結果、一致したセルを示す。又、 $---$ はセル間の $r_2$ が1であることを示す。

## 4. PPM

ここではメッシュ構造のマルチプロセッサ上に任意のPMの構成を可能とする手続きを説明する。以下では、メッシュ構造をしたマルチプロセッサそれ自体もPPMと呼ぶ。但し、各プロセッサは基本的には3.1に示す4種類のセルのいずれかであるとする。

### 4.1 アルゴリズム LAYOUT<sup>(8)</sup>

説明の便宜上、メッシュ構造のPPMに対し、図5のよう

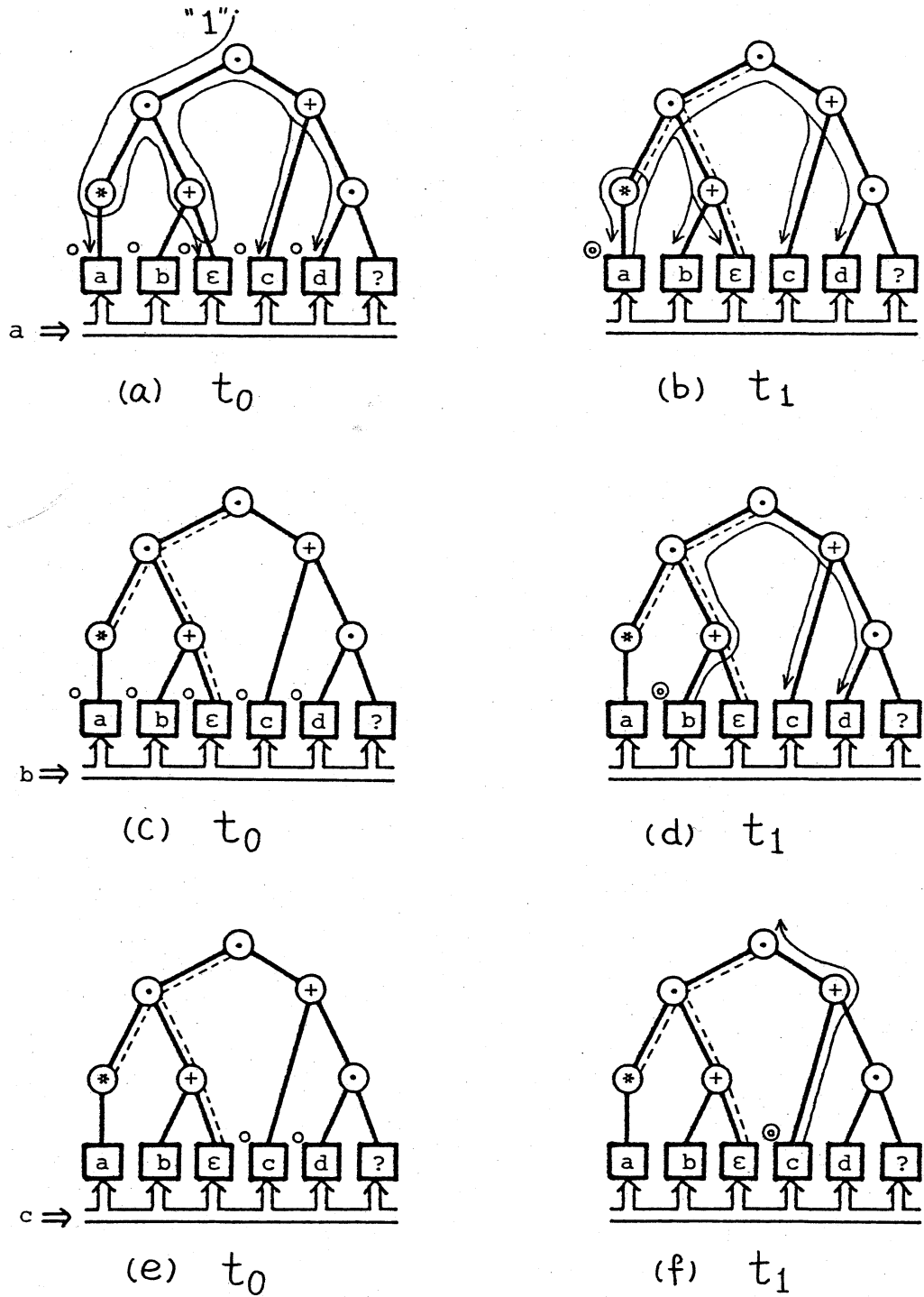


図4 PMの動作

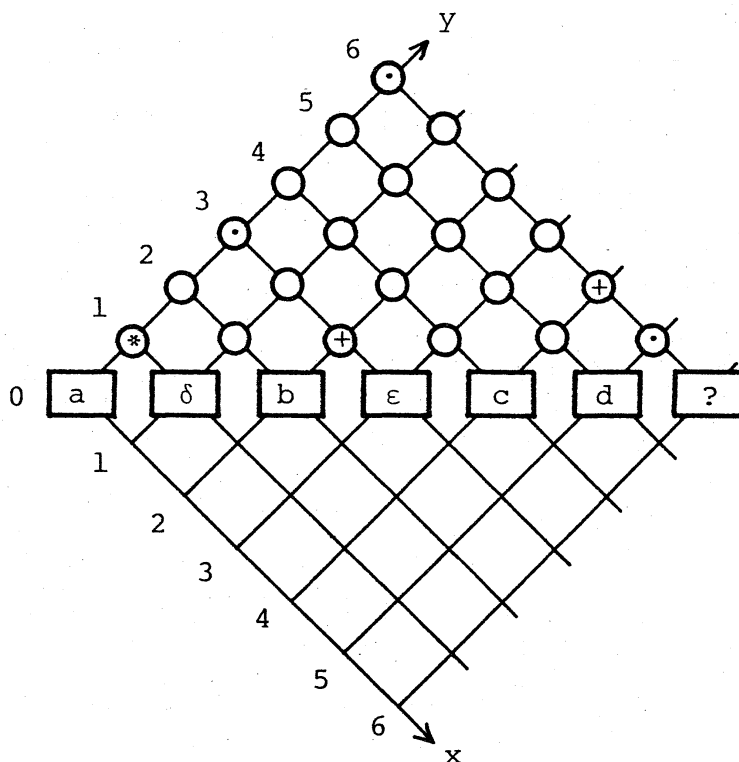


図5 LAYOUTの適用

に座標を導入し， $x$ 方向のセルの並びを「行」， $y$ 方向のセルの並びを「列」と呼ぶ．正規表現 $P$ に基づいて， $PM$ をメッシュ構造上に構成する手続きをアルゴリズム4 LAYOUTと呼ぶ．アルゴリズム4 LAYOUTを適用すると， $PM$ に含まれている演算子セルの $PPM$ 上（図5の○印）への埋込みが決定される．各セル間の結線については4.2に示すアルゴリズム4 PATHで定める．

図2と同じ例に対しアルゴリズム4 LAYOUTを適用すると図5に示すような演算子セルの埋込みが得られる．

## 4.2 アルゴリズム PATH<sup>(8)</sup>

PPM上の各演算子セルの機能が決定された後，セル間の経路を定める手続きをアルゴリズム PATH と呼ぶ。アルゴリズム PATH では図5の  $x=0, y=6$  の演算子セルから出発して，下向きに，順次，各演算子セルについてその結線を決定して行く。

図5に対し，引き続きアルゴリズム PATH を適用した結果を図6に示す。図中の点線は実際には使用されないセル間の結線を，○は信号の通過節点をそれぞれ表す。

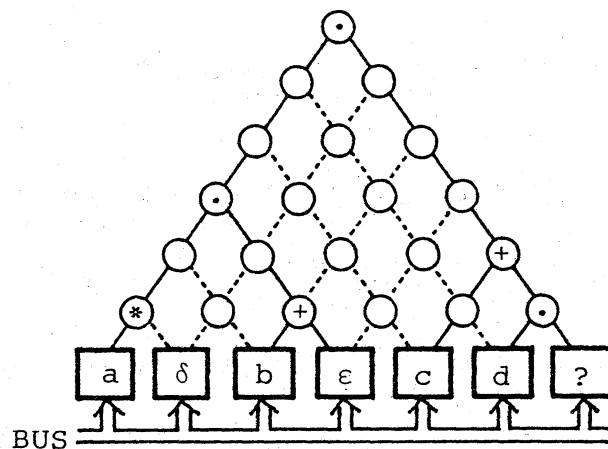


図6 PPM

上述の2つのアルゴリズム LAYOUT, PATH に関し次の命題が得られる<sup>(8)</sup>。

[命題] アルゴリズム LAYOUT, PATHにより，任意の正規表現  $P$  に対する PM がメッシュ構造の PPM 上に正しく構成



される。

最後に、日頃御指導いただく本学吉田典可教授に深く感謝いたします。

### 文献

- (1) Chu, K.H. and Fu, K.S.: "VLSI architectures for high speed recognition of context-free languages and finite-state languages", Proc. 9th Annual Int'l Symp. on Computer Architecture, pp.43-49 (1982).
- (2) Floyd, R.W. and Ullman, J.D.: "The compilation of regular expressions into integrated circuits", J.Assoc.Comput.Mach., 29, 3, pp.603-622 (1982).
- (3) Foster, M.J. and Kung, H.T.: "Recognize regular languages with programmable building-blocks", in J.P.Gray, ed., "VLSI 81", pp.75-84, Academic Press, London (1981).
- (4) Hopcroft, J.E. and Ullman, J.D.: "Introduction to Automata Theory, Languages and Computation", Addison-Wesley, Reading, MA (1979).
- (5) Mead, C. and Conway, L.: "Introduction to VLSI Systems", Addison-Wesley, Reading, MA (1980).
- (6) Mukhopadhyay, A.: "Hardware algorithms for nonnumeric computation", IEEE Trans. Comput., C-28, 6, pp.384-394 (1979).
- (7) 若林, 菊野, 吉田, 藤井: "正規集合を受理するプログラムマブルパターンマッチングマシン (PPM)", 昭和58年電子通信学会総合全国大会発表予定。

- (8) Wakabayashi, S.: "Programmable pattern matching machine for high speed recognition of regular set", Hiroshima Univ. ECS Tech. Rep. No.83-03 (1983).