



Office de la Propriété
Intellectuelle
du Canada

Un organisme
d'Industrie Canada

Canadian
Intellectual Property
Office

An agency of
Industry Canada

CA 2820500 A1 2007/06/07

(21) **2 820 500**

(12) **DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION**

(13) **A1**

(22) Date de dépôt/Filing Date: 2006/11/29

(41) Mise à la disp. pub./Open to Public Insp.: 2007/06/07

(62) Demande originale/Original Application: 2 629 618

(30) Priorité/Priority: 2005/12/02 (US11/293,619)

(51) Cl.Int./Int.Cl. *G06F 17/00* (2006.01)

(71) Demandeur/Applicant:
EXEGY INCORPORATED, US

(72) Inventeurs/Inventors:
CYTRON, RON K., US;
TAYLOR, DAVID EDWARD, US;
BRODIE, BENJAMIN CURRY, US

(74) Agent: MCFADDEN, FINCHAM INC.

(54) Titre : PROCEDE ET DISPOSITIF D'APPARIEMENT D'EXPRESSIONS NORMALES HAUTE PERFORMANCE

(54) Title: METHOD AND DEVICE FOR HIGH PERFORMANCE REGULAR EXPRESSION PATTERN MATCHING

(57) **Abrégé/Abstract:**

A device for matching an input string to a pattern via a deterministic finite automaton (DFA), the DFA comprising a plurality of states including a current state and a plurality of possible next states, the input string comprising a plurality of input symbols. The device comprises at least two parallel pipeline stages; a first one of the pipeline stages being configured to retrieve a plurality of transitions to a possible next state of the DFA from a pre-populated memory, and a second one of the pipeline stages configured to choose, based at least in part upon the DFA's current state, one of said retrieved transitions from which the integrated circuit will determine the next state of the DFA, wherein the second one of the pipeline stages is downstream from the first one of the pipeline stages.



ABSTRACT

A device for matching an input string to a pattern via a deterministic finite automaton (DFA), the DFA comprising a plurality of states including a current state and a plurality of possible next states, the input string comprising a plurality of input symbols. The device comprises at least two parallel pipeline stages; a first one of the pipeline stages being configured to retrieve a plurality of transitions to a possible next state of the DFA from a pre-populated memory; and a second one of the pipeline stages configured to choose, based at least in part upon the DFA's current state, one of said retrieved transitions from which the integrated circuit will determine the next state of the DFA, wherein the second one of the pipeline stages is downstream from the first one of the pipeline stages.

Method and Device for High Performance Regular Expression
Pattern Matching

This is a division of Canadian Serial No. 2,629,618 filed November 29, 2006.

Field of the Invention:

The present invention relates generally to the field of processing a stream of data symbols to determine whether any strings of the data symbol stream match a pattern.

5

Background and Summary of the Invention:

Advances in network and storage-subsystem design continue to push the rate at which data streams must be processed between and within computer systems. Meanwhile, the content
10 of such data streams is subjected to ever increasing scrutiny, as components at all levels mine the streams for patterns that can trigger time sensitive action. Patterns can include not only constant strings (e.g., "dog" and "cat") but also specifications that denote credit card numbers, currency
15 values, or telephone numbers to name a few. A widely-used pattern specification language is the regular expression language. Regular expressions and their implementation via deterministic finite automata (DFAs) is a well-developed field. See Hopcroft and Ullman, *Introduction to Automata*
20 *Theory, Languages, and Computation*, Addison Wesley, 1979. A DFA is a logical representation that defines the

operation of a state machine, as explained below. However, the inventors herein believe that a need in the art exists for improving the use of regular expressions in connection with high performance pattern matching.

5 For some applications, such as packet header filtering, the location of a given pattern may be anchored, wherein anchoring describes a situation where a match occurs only if the pattern begins or ends at a set of prescribed locations within the data stream. More commonly, in many applications,
 10 a pattern can begin or end anywhere within the data stream (e.g., unstructured data streams, packet payloads, etc.). Some applications require a concurrent imposition of thousands of patterns at every byte of a data stream. Examples of such applications include but are not limited to:

- 15 • network intrusion detection/prevention systems (which typically operate using a rule base of nearly 10,000 patterns (see Roesch, M., "Snort - lightweight intrusion detection for networks", LISA '99: 13th Systems Administration Conference, pp. 229-238, 1999);
- 20 • email monitoring systems which scan outgoing email for inappropriate or illegal content;
- spam filters which impose user-specific patterns to filter incoming email;
- 25 • virus scanners which filters for signatures of programs known to be harmful; and
- copyright enforcement programs which scan media files or socket streams for pirated content.

In applications such as these, the set of patterns sought within the data streams can change daily.

30 Today's conventional high-end workstations cannot keep pace with pattern matching applications given the speed of data streams originating from high speed networks and storage subsystems. To address this performance gap, the inventors herein turn to architectural innovation in the formulation and

35

realization of DFAs in pipelined architectures (e.g., hardware logic, networked processors, or other pipelined processing systems).

5 A regular expression r denotes a regular language $L(r)$, where a language is a (possibly infinite) set of (finite) strings. Each string is comprised of symbols drawn from an alphabet Σ . The syntax of a regular expression is defined inductively, with the following basic expressions:

- any symbol $\alpha \in \Sigma$ denotes $\{ \alpha \}$;
- 10 • the symbol λ denotes the singleton set containing an empty (zero-width) string; and
- the symbol ϕ denotes the empty set.

Each of the foregoing is a regular language. Regular expressions of greater complexity can be constructed using the union, concatenation, and Kleene-closure operators, as is 15 well-known in the art. Symbol-range specifiers and clause repetition factors are typically offered for syntactic convenience. While any of the well-known regular expression notations and extensions are suitable for use in the practice of the present invention, the description herein and the 20 preferred embodiment of the present invention supports the perl notation and extensions for regular expressions due to perl's popularity.

As noted above, regular expressions find practical use in 25 a plethora of searching applications including but not limited to file searching and network intrusion detection systems. Most text editors and search utilities specify search targets using some form of regular expression syntax. As an illustrative example, using perl syntax, the pattern shown in 30 Figure 1 is intended to match strings that denote US currency values:

- a backslash "\" precedes any special symbol that is to be taken literally;
- the low- and high-value of a character range is specified 35 using a dash "-";

- the "+" sign indicates that the preceding expression can be repeated one or more times;
- a single number in braces indicates that the preceding expression can be repeated exactly the designated number of times; and
- a pair of numbers in braces indicates a range of repetitions for the preceding expression.

Thus, strings that match the above expression begin with the symbol "\$", followed by some positive number of decimal digits; that string may optionally be followed by a decimal point "." and exactly two more decimal digits. In practice, a pattern for such matches may also specify that the match be surrounded by some delimiter (such as white space) so that the string "\$347.12" yields one match instead of four matches (i.e., "\$3", "\$34", "\$347", "\$347.12").

Applications that use regular expressions to specify patterns of interest typically operate as follows: Given a regular expression r and a target string t (typically the contents of some input stream such as a file), find all substrings of t in $L(r)$. The substrings are typically reported by their position within t . Thus, unless otherwise stated, it is generally intended that the pattern r is applied at every position in the target and that all matches are reported.

The simplest and most practical mechanism for recognizing patterns specified using regular expressions is the DFA, which is formally described as the 5-tuple:

$$(Q, \Sigma, q_0, \delta, A)$$

where:

- Q is a finite set of states
- Σ is an alphabet of input symbols.
- $q_0 \in Q$ is the DFA's initial state
- δ is a transition function: $Q \times \Sigma \mapsto Q$
- $A \subseteq Q$ is a set of accepting states

A DFA operates as follows. It begins in state q_0 . If the DFA is in state q , then the next input symbol a causes a transition determined by $\delta(q, a)$. If the DFA effects a transition to a state $q \in A$, then the string processed up to that point is accepted and is in the language recognized by the DFA. As an illustrative example, the regular expression of Figure 1 can be translated into the canonical DFA shown in Figure 2 using a sequence of well-known steps, including a step that makes the starting position for a match arbitrary (unanchored) (see the Hopcroft and Ullman reference cited above). For convenience, the DFA of Figure 2 uses the term "[0-9]" to denote the set $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ and uses the symbol "~" to denote all symbols of Σ not in the set $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, \$, \cdot\}$.

The construction of a DFA typically involves an intermediate step in which a nondeterministic finite automaton (NFA) is constructed. An NFA differs from a DFA in that whereas a DFA is a finite state machine that allows at most one transition for each input symbol and state, an NFA is a finite state machine that allows for more than one transition for each input symbol and state. Also, every regular language has a canonical DFA that is obtained by minimizing the number of states needed to recognize that language. Unless specified otherwise herein, it should be assumed that all automata are in canonical (deterministic) form.

However, for the purpose of pattern matching, the inventors herein believe that the DFA shown in Figure 2 is deficient in the following respects:

- Symbols not in the alphabet of the regular expression will cause the DFA to block. For pattern-matching, such symbols should be ignored by a DFA so that it can continue to search for matches. This deficiency can be overcome by completing the DFA as follows:
 - The alphabet is widened to include any symbol that might occur in the target string. In this

description, it is assumed that Σ is the ASCII character set comprising 256 symbols.

- The DFA is augmented with a new state U :

$$Q \leftarrow Q \cup \{ U \}$$

- 5 - The transition function δ is completed by defining δ
 $(q, a) = U$ for all $q \in Q$, $a \in \Sigma$ for which δ was
previously undefined.

- A match will be found only if it originates at the first character of the target string. Pattern-matching applications are concerned with finding all occurrences of the denoted pattern at any position in the target. This deficiency can be overcome by allowing the DFA to restart at every position. Formally, a λ transition is inserted from every $q \in Q$ to q_0 .

15 The result of the above augmentation is an NFA that can be transformed into a canonical DFA through known techniques to obtain the DFA. Figure 3 provides an illustrative example of such a canonical DFA.

- 20 A DFA is typically implemented interpretively by realizing its transitions δ as a table: each row corresponds to a state of the DFA and each column corresponds to an input symbol. The transition table for the DFA of Figure 3 is shown in Figure 4. If the alphabet Σ for the DFA is the ASCII character set (as is often the case in many applications), then the transition table of Figure 4 would have 256 columns. Each entry in the transition table of Figure 4 comprises a next state identifier. The transition table of Figure 4 works thusly: if the DFA's current state is B and the next input symbol is 2, then the transition table calls for a transition to state D as "D" is the next state identifier that is indexed by current state B and input symbol 2. In the description herein, states are labeled by letters to avoid confusion with symbol encodings. However, it is worth noting that in

practice, states are typically represented by an integer index in the transition table.

The inventors herein believe that the pattern matching techniques for implementing DFAs in a pipelined architecture
5 can be greatly improved via the novel pattern matching architecture disclosed herein. According to one aspect of the present invention, a pipelining strategy is disclosed that defers all state-dependent (iterative, feedback dependent) operations to the final stage of the pipeline. Preferably,
10 transition table lookups operate to retrieve all transition table entries that correspond to the input symbol(s) being processed by the DFA. Retrievals of transition entries from a transition table memory will not be based on the current state of the DFA. Instead, retrievals from the transition table
15 memory will operate to retrieve a set of stored transition entries based on data corresponding to the input symbol(s) being processed.

In a preferred embodiment where alphabet encoding is used to map the input symbols of the input data stream to
20 equivalence class identifiers (ECIs), these transition table entries are indirectly indexed to one or more input symbols by data corresponding to ECIs. This improvement allows for the performance of single-cycle state transition decisions, enables the use of more complex compression and encoding
25 techniques, and increases the throughput and scalability of the architecture.

According to another aspect of the present invention, the transitions of the transition table preferably include a match flag that indicates whether a match of an input symbol string
30 to the pattern has occurred upon receipt of the input symbol(s) that caused the transition. Similarly, the transitions of the transition table preferably include a match restart flag that indicates whether the matching process has restarted upon receipt of the input symbol(s) that caused the
35 transition. The presence of a match flag in each transition allows for the number of states in the DFA to be reduced

relative to traditional DFAs because the accepting states can be eliminated and rolled into the match flags of the transitions. The presence of a match restart flag allows the DFA to identify the substring of the input stream that matches an unanchored pattern. Together, the presence of these flags in the transitions contribute to another aspect of the present invention - wherein the preferred DFA is configured with an ability to scale upward in the number of bytes processed per cycle. State transitions can be triggered by a sequence of m input symbols, wherein m is greater than or equal to 1 (rather than being limited to processing only a single input symbol per clock cycle). Because of the manner by which the transitions include match flags and match restart flags, as disclosed herein, the DFA will still be able to detect when and where matches occur in the input stream as a result of the leading or an intermediate input symbol of the sequence of m input symbols that are processed together by the DFA as a group.

According to yet another aspect of the present invention, incremental scaling, compression and character-encoding techniques are used to substantially reduce the resources required to realize a high throughput DFA. For example, run-length coding can be used to reduce the amount of memory consumed by (i.e., compress) the DFA's transition table. Furthermore, the state selection logic can then operate on the run-length coded transitions to determine the next state for the DFA. Masking can be used in the state selection logic to remove from consideration portions of the transition table memory words that do not contain transitions that correspond to the ECI of the input symbol(s) being processed.

Also, according to yet another aspect of the present invention, a layer of indirection can be used to map ECIs to transitions in the transition table memory. This layer of indirection allows for the use of various optimization techniques that are effective to optimize the run-length coding process for the transition entries in the transition

table memory and optimize the process of effectively packing the run-length coded transition entries into words of the transition table memory such that the number of necessary accesses to transition table memory can be minimized. With
5 the use of indirection, the indirection entries in the indirection table memory can be populated to configure the mappings of ECIs to transition entries in the transition table memory such that those mappings take into consideration any optimization processes that were performed on the transition
10 entries in the transition table memory.

Furthermore, according to another aspect of the present invention, disclosed herein is an optimization algorithm for ordering the DFA states in the transition table, thereby improving the DFA's memory requirements by increasing the
15 efficiency of the run-length coded transitions.

Further still, disclosed herein is an optimization algorithm for efficiently packing the transition table entries into memory words such that the number of transition table entries sharing a common corresponding input symbol (or
20 derivative thereof such as ECI) that span multiple memory words is minimized. This memory packing process operates to improve the DFA's throughput because the efficient packing of memory can reduce the number of memory accesses that are needed when processing one or more input symbols.

25 According to another aspect of the present invention, the patterns applied during a search can be changed dynamically without altering the logic of the pipeline architecture itself. A regular expression compiler need only populate the transition table memory, indirection table, ECI mapping
30 tables, and related registers to reprogram the pattern matching pipeline to a new regular expression.

Based on the improvements to DFA design presented herein, the inventors herein believe that the throughput and density achieved by the preferred embodiment of the present invention
35 greatly exceed other known pattern matching solutions.

These and other inventive features of the present invention are described hereinafter and will be apparent to those having ordinary skill in the art upon a review of the following specification and figures.

5

Brief Description of the Drawings:

Figure 1 depicts an exemplary regular expression;

Figure 2 depicts a DFA for the regular expression of Figure 1;

10 Figure 3 depicts an improved DFA for the regular expression of Figure 1;

Figure 4 depicts a conventional transition table for the DFA of Figure 3;

15 Figure 5 illustrates a block diagram overview of a preferred embodiment of the present invention;

Figure 6 depicts a preferred algorithm for doubling the stride of a DFA;

20 Figure 7 depicts a DFA in accordance with the preferred embodiment of the present invention having a reduced number of states and flags within the transition for matches and match restarts;

Figure 8 depicts a preferred algorithm for encoding input symbols into equivalence class identifiers (ECIs);

25 Figures 9(a) and (b) depict a preferred transition table for the DFA of Figure 7;

Figures 10(a) and (b) depict transition tables for the regular expression of Figure 1 wherein the stride of the DFA is equal to 2 input symbols per cycle;

30 Figure 11 depicts a preferred algorithm for constructing a base DFA¹ from d and p;

Figure 12 depicts an exemplary run-length coded transition table;

Figure 13 depicts an adjacently-stored version of the run-length coded transition table of Figure 12;

35 Figure 14 depicts an exemplary state selection logic circuit for determining a next state based on a retrieved set

of run-length coded transitions that correspond to a given ECI;

Figures 15(a) and (b) depict an indirection table and a memory in which the run-length coded transition table of
5 Figures 12 and 13 is stored;

Figures 16(a) and (b) depict an indirection table and a memory in which the run-length coded transition table of Figure 15 include precomputed run-length prefix sums;

Figures 17(a) and (b) depict an alternative formulation
10 of the Indirection Table and TTM in which the run-length coded transition table of Figure 14 include precomputed run-length prefix sums;

Figure 18 illustrates an exemplary state selection logic circuit;

15 Figures 19(a) and (b) respectively illustrate an exemplary transition table that has been optimized by state re-ordering, and the run length-coded version of the state re-ordered transition table;

Figure 20 depicts a preferred algorithm for calculating
20 the difference matrix used to optimize the run-length coding for the TTM;

Figure 21 depicts a preferred algorithm for optimization of run-length coding in the TTM;

Figure 22 depicts an example of a coded transition table
25 that has been packed into the TTM;

Figure 23 depicts a preferred algorithm for optimal packing of memory words;

Figure 24 depicts a preferred regular expression system architecture;

30 Figure 25 depicts the regular expression engine of the preferred embodiment as a series of pipelined computational blocks; and

Figure 26 depicts a preferred process for assigning ECIs to input symbols using direct-address tables and pairwise
35 combination.

Detailed Description of the Preferred Embodiments:

Figure 5 depicts an overview of the preferred embodiment of the present invention. The architecture of the preferred embodiment is illustrated within the regular expression circuit 502, which serves as a pattern matching circuit that operates on an input data stream comprising a plurality of sequential input symbols. Preferably, the regular expression circuit 502 is implemented in hardware logic (e.g., reconfigurable hardware logic such as an FPGA or nonreconfigurable hardware logic such as an ASIC). It is worth noting that one or more of the regular expression circuits 502i can be implemented on the same device if desired by a practitioner of the present invention, which is also reflected in Figure 24. Also, the regular expression circuit can be implemented in other pipelined architectures, such as multi-processor systems, wherein each processor would serve as a pipeline stage of the regular expression circuit. In such an example, the different processors of the pipeline can be networked together.

The data tables and relevant registers of the regular expression circuit are preferably populated by the output of the regular expression compiler 500. Regular expression compiler 500 operates to process a specified (preferably user-specified) regular expression to generate the DFA that is realized by the regular expression circuit 502 as described herein. Preferably, regular expression compiler 500 is implemented in software executed by a general purpose processor such as the CPU of a personal computer, workstation, or server.

Regular expression compiler 500 can be in communication with regular expression circuit 502 via any suitable data communication technique including but not limited to networked data communication, a direct interface, and a system bus.

The regular expression circuit 502 preferably realizes the DFA defined by one or more specified regular expressions via a plurality of pipelined stages. A first pipeline stage

is preferably an alphabet encoding stage 504 that produces an ECI output from an input of m input symbols, wherein m can be an integer that is greater than or equal to one. A second pipeline stage is preferably an indirection table memory stage 506. The indirection table memory stage 506 can be addressed in a variety of ways. Preferably, it is directly addressed by the ECI output of the alphabet encoding stage 504. A third pipeline stage is the transition table logic stage 508 that operates to receive an indirection table entry from the output of the indirection table memory stage 506 and resolve the received indirection entry to one or more addresses in the transition table memory stage 510. The transition table logic stage 508 also preferably resolves the received indirection table entry to data used by the state selection logic stage 512 when the stage selection logic stage processes the output from the transition table memory stage 510 (as described below in connection with the masking operations).

The transition table memory stage 510 stores the transitions that are used by the DFA to determine the DFA's next state and determine whether a match has been found. The state selection logic stage 512 operates to receive one or more of the transition entries that are output from the from the transition table memory stage 510 and determine a next state for the DFA based on the DFA's current state and the received transition(s). Optionally, the masking operations 514 and 516 within the state selection logic stage 512 that are described below can be segmented into a separate masking pipeline stage or two separate masking pipeline stages (an initial masking pipeline stage and a terminal masking pipeline stage). Additional details about each of these stages is presented herein.

High-Throughput DFAs

A conventional DFA processes one input symbol (byte) at a time, performing a table lookup on each byte to determine the next state. However, modern communication interfaces and

interconnects often transport multiple bytes per cycle, which makes the conventional DFA a "bottleneck" in terms of achieving higher throughput. Throughput refers to the rate at which a data stream can be processed - the number of bytes per second that can be accommodated by a design and its implementation.

An extension of conventional DFAs is a DFA that allows for the performance of a single transition based on a string of m symbols. See Clark and Schimmel, "Scalable pattern matching for high speed networks", IEEE Symposium on Field-Programmable Custom Computing Machines, April 2004. That is, the DFA processes the input stream in groups of m input symbols. Formally, this adaptation yields a DFA based on the alphabet Σ^m ; the corresponding transition table is of size $|Q| |\Sigma|^m$. This apparently dramatic increase in resource requirements is mitigated by the compression techniques described herein. For convenience, we let δ^m denote a transition function that operates on sequences of length m , with $\delta = \delta^1$.

As an illustrative example, consider doubling the effective throughput of the DFA shown in Figure 3 by processing two bytes at a time ($m = 2$). Based on the table in Figure 4, if the current state is E, then the input sequence "2\$" would result in a transition to state B: $\delta^2(E, 2\$) = \delta(\delta(E, 2), \$) = B$. By accounting for all such two-character sequences, a complete transition table can be computed for this higher-throughput DFA as explained below.

In general, an algorithm for constructing δ^m for a given DFA is straightforward. The set of states is unchanged and the transition function (table) is computed by simulating progress from each state for every possible sequence of length m . That algorithm takes time $\theta(|Q| |\Sigma|^m m)$ to compute a table of size $\theta(|Q| |\Sigma|^m)$. A faster algorithm can be obtained by the following form of dynamic programming. Consider

$$m = 2^i, i > 0 \text{ and a string } x = x_l x_r, |x_l| = |x_r| = \frac{m}{2}$$

Then

$$\forall q \delta^m(q, x) = \delta^{\frac{m}{2}}(\delta^{\frac{m}{2}}(q, x_l), x_r)$$

5 An algorithm based on the above proposition is shown in Figure 6.

To obtain higher throughput DFAs, the algorithm in Figure 6 can be invoked repeatedly, each time doubling the effective number of symbols processed per cycle. This reduces the complexity of computing δ^m to $O(|Q| |\Sigma|^m \log m)$. Moreover, by
10 identifying redundant columns, the effective alphabet of each table can be substantially reduced in practice through encoding, as described in the Alphabet Encoding section below.

High-Throughput DFAs: Accepts

15 Because the higher throughput DFA performs multiple transitions per cycle, it can traverse an accepting state of the original DFA during a transition. We therefore augment the transition function to include whether an accepting state is traversed in the trace:

$$20 \quad \delta^m : Q \times \Sigma^m \rightarrow Q \times \{0, 1\}$$

The range's second component indicates whether the sequence of symbols that caused the transition contains a nonempty prefix that takes the original DFA through an accept state.

Transition functions of this form obviate the need for a
25 set of accepting states A , because the "accept" (match) information is associated with edges of the higher throughput DFA. This is formalized via the modified DFA we define in the "Synergistic Combination of Stride and Encoding" section below.

30 For $m > 1$, accepts are now imprecise because the preferred DFA does not keep track of which intermediate symbol actually caused an accept (match) to occur. To favor speed, the high-throughput DFA can be configured to allow imprecise

accepts, relegating precise determination of the accept point to software postprocessing.

High-Throughput DFAs: Restarts

5 As previously discussed, a pattern-matching DFA for a regular expression is preferably augmented with transitions that allow matches to occur throughout the target string. Because matches can occur at any starting position in the target, an accept should report the origination point of the match in the target. It is not clear in the automaton of Figure 3 when the origination point is set. For example, all transitions to state A set the origination point, but so do transitions on "\$" to state B. Considering transitions from E to A, a "~" or "." is a restart, while a digit is an accept.

15 The λ -transitions introduced to achieve position independence of matching result in an NFA that can be transformed into a DFA through the usual construction. The "Synergistic Combination of Stride and Encoding" section below describes how to modify that construction to identify transitions that serve only to restart the automaton's matching process.

Formally, the transition function is augmented once more, this time to indicate when restarts occur:

$$\delta' : Q \times \Sigma^m \rightarrow Q \times \{0, 1\} \times \{0, 1\}$$

25 The first flag indicates a restart transition (a "match restart" flag) and the second flag indicates an accept transition (a "match" flag). Accordingly, the DFA diagrams henceforth show restart transitions with green edges and accept edges with red edges. For example, Figure 7 shows an illustrative example of a diagram for an automaton that processes one symbol at a time and recognizes the language denoted by Figure 1. Optionally, the edges for the transitions to State A can be coded as black and flagged accordingly, with only the edges for the transitions to State B being coded as green and flagged accordingly. The actions of a DFA with the colored edges are as follows. The automaton

includes context variables b and e to record the beginning and end of a match; initially, $b = e = 0$, and the index of the first symbol of the target is 1. These variables allow the location of a matching string to be found in the context
 5 buffer as shown in Figure 24. Transitions are then performed as follows:

black: e is advanced by m —the stride of the automaton, which is the length of the input string that caused the
 10 transition. In Figure 6, $m = 1$. A match is in progress and the portion of the target string participating thus far begins at position b and ends at position e , inclusively.

red only: e is advanced by m and a match is declared. The
 15 target substring causing the match starts at position b and ends at position e .

green only: b is advanced by m and e is set to b . The automaton is restarting the match process.
 20

red and green: The red action is performed before the green action.

Figure 9 shows the associated transition table for the
 25 DFA of Figure 7, wherein each transition entry in the transition table includes a match flag and a match restart flag in addition to the next state identifier. Because information about accepts is associated with edges (the δ function), a DFA with colored edges can have fewer states than
 30 the canonical DFA.

The use of match flags and match restart flags is particularly useful when scaling the DFA to process multiple input symbols per cycle. Figure 10(b) illustrates a transition table for the regular expression of Figure 1
 35 wherein m is equal to 2. Figure 10(a) depicts the symbol encoding for this example. Thus, even when processing

multiple input symbols per cycle, the DFA will be able to detect when matches and restarts occur on the leading or an intermediate symbol of the group m of input symbols because the flags will carry match status information rather than the
 5 states.

Alphabet Encoding

As described above, the size of the transition table (δ) increases exponentially with the length of the input sequence
 10 consumed in each cycle. In this section, techniques are presented to encode the symbol alphabet, the goal of which is to mitigate the transition table's size and thus maximize the number of symbols processed per cycle.

Frequently, the set of symbols used in a given regular
 15 expression is small compared with the alphabet Σ of the search target. Symbols present in the target but not denoted in the pattern will necessarily be given the same treatment in the DFA for the regular expression. More generally, it may be the case that the DFA's behavior on some set of symbols is
 20 identical for all symbols in that set. As an illustrative example, the regular expression in Figure 1 uses only a small fraction of the ASCII character set. The transitions for digits are the same in the DFA shown in Figure 3, and the symbol "~" stands for all symbols of the ASCII set that are
 25 not in the set $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, \$, .\}$.

While a regular expression may mention character classes explicitly, such as "[0 - 9]", a more general approach is achieved by analyzing a DFA for equivalent state-transition behavior. Formally, if

30 $(\exists a \in \Sigma) (\exists b \in \Sigma) (\forall q \in Q) \delta(q, a) = \delta(q, b)$
 then it can be said that a and b are "transition equivalent."

Given a transition table $\delta : Q \times \Sigma \mapsto Q$, an
 $O(|\Sigma|^2|Q|)$ algorithm for partitioning Σ into equivalence
 classes is shown in Figure 8. Using the example of Figure 1,
 35 the algorithm develops the equivalence classes of symbols suggested in Figure 1 to form the 4 columns shown in Figure

9(b). The partition is represented by a set of integers K and a mapping k from Σ to K . Because alphabet symbol-equivalence is determined by state equivalence, the best result is obtained if δ corresponds to a canonical DFA, with state
 5 equivalence already determined.

Computing equivalence classes using the DFA, rather than inspection of its associated regular expression, is preferred for the following reasons:

- The regular expression may be specified without using
 10 ranges or other gestures that may imply equivalence classes. As an illustrative example, a and b can be made equivalent in the DFA for the regular expression " $(ac)|(bc)$ ", but their equivalence is not manifest in the regular expression.
- Equivalence classes often cannot be determined by local
 15 inspection of a regular expression. As an illustrative example, the regular expression " $[0-9]a|5c$ " contains the phrase " $[0-9]$ ", but one element of that range (5) will not share identical transitions with the other symbols
 20 because of the other phrase. The appropriate partition in this case is $\{ \{ 0, 1, 2, 3, 4, 6, 7, 8, 9 \}, \{ 5 \} \}$

Formally, the function in Figure 8 creates a partition — a set, each element of which is a set of symbols in Σ that are
 25 treated identically in the DFA. The function produces a more convenient form of that partition in k , which maps each symbol to an integer denoting its equivalence class (the integer serving as an "equivalence class identifier" (ECI)). A DFA based on symbol encodings then operates in the following two
 30 stages. First, the next input symbol is mapped by k to its ECI that represents its equivalence class as determined by the algorithm in Figure 8. Second, the DFA can use its current state and the ECI to determine the DFA's next state.

Based on the ideas presented thus far, Figure 7 shows an
 35 illustrative example of a DFA that processes a single symbol at a time, with red edges indicating "accept" and green edges

indicating "restart". Analysis using the algorithm in Figure 8 yields the symbol encoding shown in Figure 9(a) and the transition table shown in Figure 9(b). As shown in Figure 9(b), each entry in the transition table is indexed by data
 5 corresponding to the DFA's state and by data corresponding to the ECI for a group of m input symbols. Figures 10(a) and (b) show the counterpart ECI table and transition table for the regular expression of Figure 1 where 2 input symbols are processed at a time.

10 Each transition in the transition table is a 3-tuple that comprises a next state identifier, a match restart flag and a match flag. For example, the transition indexed by state D and ECI 0 is (B, 1, 0) wherein B is the next state identifier, wherein 1 is the match restart flag, and wherein 0 is the
 15 match flag. Thus, the transition from state D to B caused by ECI 0 can be interpreted such that ECI 0 did not cause a match to occur but did cause the matching process to restart.

Synergistic Combination of Stride and Encoding

20 The ideas of improving throughput and alphabet encoding discussed above are now combined to arrive at an algorithm that consumes multiple bytes per cycle and encodes its input to save time (in constructing the tables) and space (in realizing the tables at runtime).

25 Such a new high-throughput DFA^m can now be formally described as the 6-tuple $(Q, \Sigma, q_0, K, \kappa, \delta)$ where:

- Q is a finite set of states
- Σ is an alphabet of the target's input symbols
- $q_0 \in Q$ is an initial state
- 30 • K is a set of integers of size $O(|\Sigma|^m)$ (but expected to be small in practice)
- κ is a function $\Sigma^m \mapsto K$ that maps m input symbols at a time to their encoding

- δ is a function $Q \times K \mapsto Q \times \{0, 1\} \times \{0, 1\}$ that maps the current state and next substring of m symbols to a next state, a possible restart, and a possible accept.

5 The set of transformations begins with a regular expression r and perform the following steps to obtain DFA^m :

1. A DFA d is constructed for one or more regular expressions r in the usual way (see the Hopcroft and
10 Ullman reference cited above). For example, as discussed above, the regular expression in Figure 1 results in the DFA shown in Figure 3.

2. An set of transitions p is computed that would allow the
15 automaton to accept based on starting at any position in the target. This is accomplished by simulating for each state a λ -transition to q_0 . Specifically, p is computed as follows:

```

     $p \leftarrow \emptyset$ 
20   foreach  $\phi \in Q$  do
       foreach  $a \in \Sigma$  do  $p \leftarrow p \cup \{ (p, a) \mapsto \delta(q_0, a) \}$ 
    }
```

3. From d and p , the base DFA^1 is constructed by the
25 algorithm in Figure 11.

4. State minimization is performed on a high-throughput DFA
by a standard algorithm (see the Hopcroft and Ullman
reference cited above), except that states are initially
30 split by incoming edge color (black, green, red), instead of by whether they are final or nonfinal states in a traditional DFA.

5. Given a DFA^k , a higher throughput DFA^{2k} with alphabet
35 encoding is constructed by the algorithm shown in Figure 6.

Transition Table Compression via Run-Length Coding

The transition table for the preferred high-throughput DFA may contain $|K| \times |Q|$ entries. State minimization
 5 attempts to minimize $|Q|$ and the previous discussion regarding the combination of higher throughput and alphabet encoding attempts to minimize $|K|$. Nonetheless, storage resources are typically limited; therefore, a technique for accommodating as many tables as possible should be addressed. The following
 10 addresses this matter by explaining how to compress the table itself.

Based on the discussion above, a transition table cell contains the three-tuple: (next state, start flag, accept flag). Run-length coding is a simple technique that can
 15 reduce the storage requirements for a sequence of symbols that exhibits sufficient redundancy. The idea is to code the string a^n as the run-length n and the symbol a ; the notation $n(a)$ can be used. Thus, the string `aaaabbbcbbaaa` is run-length coded as `4(a)3(b)1(c)2(b)3(a)`. If each symbol and each
 20 run-length requires one byte of storage, then run-length coding reduces the storage requirements for this example by three bytes (from 13 bytes to 10 bytes).

Examining the example of Figure 9(b), there is ample opportunity for run-length coding in the columns of the
 25 transition table. For the encoded ECI symbols 0 and 3, the table specifies the same three-tuple for every previous state, so the run-length coding prefix for the transitions in the table indexed by ECIs 0 and 3 are both "5". In general, what is expected in transition tables is common "next-state"
 30 behavior. Thus, the number of unique entries in each column of the transition table is typically smaller than the number of states in the DFA. Figure 12 contains a run-length coded version of the transition table in Figure 9.

While column compression can save storage, it appears to
 35 increase the cost of accessing the transition table to obtain a desired entry. Prior to compression, a row is indexed by

the current state, and the column is indexed by the ECI. Once the columns are run-length coded, as shown in Figure 12, the compressed contents of each column are stored adjacently, as shown in Figure 13. In this example, the states themselves
 5 have also been encoded as integers (with A represented by 0, B by 1, etc.). There are now two steps to determining the DFA's next state and actions:

1. Based on the next ECI encoded from the input, the relevant range of entries is found in the storage layout
 10 shown in Figure 13. This lookup can be performed using a mapping from ECI to offset in the storage layout. The range of entries is a compressed column from Figure 12. In Figure 13, the underscored entries correspond to ECI 1 from Figure 12.
- 15 2. Based on the current state, the next state and action flags must be found in the relevant range of entries. This logic, called *state selection*, essentially requires decompressing the entries to discover which entry corresponds to the index of the current state.

20 If an entire compressed column is available, the circuit shown in Figure 14 depicts an example of how state selection can be realized by interpreting the compressed form and the current state to determine the appropriate tuple entry. For
 25 each entry in the run-length coded column, the sum of its run-length prefix and the run-length prefixes of the preceding entries are computed. The current state index is then compared with each sum and the first (leftmost in the example of Figure 14) entry whose sum is greater than the current
 30 state index is selected by the priority encoder. The priority encoder can then determine the next state for the DFA.

Figure 14 shows an illustrative example of state selection in progress for ECI 1 of Figure 12. Each "<" node compares the current state index (3 in this example, or State
 35 D) with the sum of run-length prefixes in the compressed column. If state index 0 (State A) were supplied as the

current state, all three comparators would output "1" and the priority encoder would pick the leftmost one, choosing (A,1,0) as the contents. In Figure 14, the current state is 3, which is less than the second sum ($3 + 1$) and the third sum ($3 + 1 + 1$), so that the right two comparators output "1". The priority encoder picks the leftmost one, so that (C,0,0) is chosen as the lookup of ECI 1 and state 3.

Supporting Variable-Length Columns in Memory

The storage layout shown in Figure 13 must be mapped to a physical memory, in which the entire table will typically be too large to be fetched in one memory access. Field-Programmable Gate Arrays (FPGAs) and similar devices often support memory banks that can be configured in terms of their size and word-length. Moreover, the size of a given entry depends on the number of bits allocated for each field (run-length, next state identifier, match restart flag, match flag). The analysis below is based on the general assumption that x transition table entries may be retrieved per cycle.

In single port memory, this means that x will match the number of transition entries per word. For multi-port memory, this means that x will match the number of ports multiplied by the number of transition entries per word. As an example, a physical memory that supports 5 accesses per cycle and holds 3 entries per word is accommodated in the preferred embodiment by setting $x = 5 \times 3 = 15$. However, a physical memory that supports only one access per cycle and holds three entries per word is accommodated in the preferred embodiment by setting $x = 3$.

Some architectures offer more flexibility than others with respect to the possible choices for x . For example, the bits of an FPGA Block Ram can sometimes be configured in terms of the number of words and the length of each word. The following considerations generally apply to the best choice for x :

- Memory accesses are generally reduced by driving x as high as possible.
- The logic in Figure 14 grows with the number of entries that must be processed at one time. The impact of that growth on the circuit's overall size depends on the target platform and implementation. Significant FPGA resources are required to realize the logic in Figure 14.

Supporting Variable-Length Columns in Memory: Transition Table Memory

Once x is chosen, the compressed columns will be placed in the physical memory as compactly as possible. Figure 15(b) shows an example where the columns of the transition table are packed into a memory with $x = 3$. Each word in the memory is indexed by a memory address. For example, the word indexed by memory address 0 includes the following transitions 5(B, 1, 0), 3(A, 1, 0), and 1(C, 0, 0). Due to the varying length of each column, a given column may start at any entry within a row.

By introducing a layer of indirection in the transition table, it is possible to leverage the memory efficiency provided by run-length coding and compact deployment of entries in the transition table memory (TTM). Figure 15(a) shows an example of such an Indirection Table which contains one entry for each ECI. Since ECIs may be assigned contiguously, the Indirection Table may be directly addressed using the ECI value for a given input string. Each Indirection Table entry may consist of the following:

- *pointer*: address of the memory word in the Transition Table Memory containing the first entry of the run-length coded transition table column;
- *transition index*: index of the first entry of the run-length coded transition table column in the first memory word for the column;

- *transition count*: (or "count" in shorthand) the number of entries in the run-length coded transition table column;

5 Once the Indirection Table entry is retrieved using the input symbol ECI, the pointer in the retrieved entry is used to read the first memory word from the TTM. Recall x is the number of entries per memory word in the TTM. An entire column is accessed by starting at address transition index and
10 reading w consecutive words from the TTM, where w is given by:

$$w = \left\lceil \frac{\text{transition.count} + \text{transition.index}}{x} \right\rceil$$

(1)

The *transition index* and *transition count* values determine which entries in the first and last memory words participate
15 in the column. In the example in Figure 15, each TTM word is capable of storing three entries, where an entry is a run-length coded transition tuple. As can be seen in Figure 15 by virtue of shading, it can be seen that two reads of the Transition Table memory are required to fetch the column for
20 ECI 1. The particular values of *transition index* and *transition count* for ECI 1 indicate that the column begins in the second entry of the first word fetched, and continues until the first entry of the last word fetched. If TTM entries were wasted by arranging for each column to start at
25 index 0, then the number of accesses can be reduced to

$$\left\lceil \frac{\text{transition.count}}{x} \right\rceil. \text{ Because } 0 \leq \text{transition index} < x, \text{ compact}$$

storage in the TTM increases the number of accesses by at most 1.

30 As discussed below and shown in Figure 25, accesses to the Indirection Table and TTM can be pipelined with each other and with the other components of the design of the present invention. If multi-port memory is available, both tables may be stored in the same physical memory without degrading performance. However, the layer of indirection results in a

variable number of accesses to the Transition Table Memory per state transition, depending on the distribution of a run-length coded column's contents in the TTM. For a particular ECI, the number of memory accesses to retrieve the column from memory cannot exceed the number required in the direct-addressed approach. On average, the number of memory accesses per state transition is considerably less. It is believed by the inventors generally that the memory efficiency achieved via run-length coding and indirection more than compensates for the overhead incurred by storing the Indirection Table and the additional run-length field.

Furthermore, the allocation of memory for the Indirection Table is relatively straightforward, as each entry is the same size and the number of entries is equal to the maximum number of input symbol equivalence classes.

Supporting Variable-Length Columns in Memory: State Selection

The implementation of a State Select logic circuit that preferably takes into account the efficient storage layout of the TTM and offers other optimizations is now described. While the TTM offers compact storage of the compressed columns, state selection logic becomes more complex. The logic shown in Figure 14 assumes that a compressed column can be presented to the logic at once, with no extraneous entries. That logic is suboptimal for performing state selection using the TTM for the following reasons:

- A compressed column may span multiple words of the TTM.
- The start of a compressed column may begin in the middle of a TTM word. Thus, entries before the start must be suppressed for state selection.
- The end of a compressed column may occur before the end of a TTM word. Thus, entries after the end must be suppressed for state selection.

The logic shown in Figure 14 uses adders to accumulate the sum of all run lengths before each entry. Because the run

lengths are fixed inside each entry, the adders can be obviated by precomputing the prefix sums and storing them, instead of the run-lengths themselves, as the "coefficient" in each tuple. By precomputing sums, the tables shown in Figure 5 15 are transformed into the tables shown in Figure 16.

The amount of logic used to determine the beginning and end of the compressed column can also be reduced. The start of each column is specified in the Indirection table using the pointer and transition index fields, which provide the TTM 10 word containing the first entry and the index within that word of the entry. The number of words w occupied by the compressed column is then given by Equation (1). Each fully occupied word contains x entries of the compressed column. In the last word, the largest index occupied by the compressed 15 column is given by:

$$(count + index - 1) \bmod x \quad (2)$$

Logic could be deployed in the State Select circuit to compute Equation 2. However, x is a design-time parameter. By appropriate parameterization of Hardware Definition 20 Language (HDL) code, Equation 2 can be computed when the Indirection and TTM tables are generated.

Thus, the amount of computational logic can be reduced by storing the following variables for each entry in the Indirection Table:

- 25 • **Pointer:** the address of the TTM word containing the first entry in the transition table column
- **Initial Transition Index:** the index of the first entry (of the transition table column) in the first TTM word spanned by the transition table column
- 30 • **Terminal Transition Index:** the index of the last entry (of the transition table column) in the last TTM word spanned by the transition table column
- **[Additional] Word Count:** $w - 1$ where w is computed by Equation 1.

35 Continuing this example, Figure 17 shows the Indirection Table and TTM entries for the transition table shown in Figure 16.

Essentially, the transition count values are translated into the word count values and terminal transition index values. This translation does not affect the contents of the TTM, but reduces the logic needed to process these tables.

5 The State Select block logic that is shown in Figure 18 operates on a transition table column containing two entries that spans one TTM word. Each word stores four entries. The output of the Word Counter shown in Figure 18 reflects the number of memory words that have been examined for the current
10 transition table row. If the current memory word is the first word spanned by the row, then the Initial Transition Index is used to retrieve a set of mask bits from the Initial Transition Index Mask ROM. These bits are used to mask off preceding entries in the memory word that are not part of the
15 transition table column. Masking is accomplished by forcing the run-length sums to be zero. Note that if the current memory word is not the first word spanned by the column, then no entries are masked at this stage.

 The next stage "masks" the entries in the last memory
20 word that are not part of the transition table column. The run-length sums for entries that are not part of the transition table column are forced to the value of the Max Run-Length Register. This value records the maximum number of entries in a transition table column (i.e. the number of
25 columns in the uncoded transition table; also the value of the run-length sum for the last entry in each coded transition table column). If the current memory word is the last memory word spanned by the transition table column (value of the Word Counter is equal to Word Count), then the Terminal Transition
30 Index is used as the address to the Terminal Transition Index Mask ROM. If this is not the case, then no entries are masked during this stage. Forcing the run-length sums of trailing entries to be the maximum run-length sum value simplifies the Priority Encoder that generates the select bits for the
35 multiplexer that selects the next state. This masking process produces an output vector from the less-than comparisons with

the following property: the index of the left-most '1' bit is the index of the next state entry, and all bits to right of this bit will be set to '1'. As previously referenced, it should be noted that the masking stages may be pipelined to
 5 increase throughput. In an alternative embodiment, only the less than comparisons, priority encoder, and next state selection logic need to occur in the final pipeline stage.

Optimizations

10 Achieving a high-throughput regular expression pattern-matching engine is the primary motivation for developing the high-throughput DFA, character encoding, and transition table compression techniques that are disclosed herein. In the following, techniques that optimize the throughput of the
 15 system at the expense of some memory efficiency are examined; thus, each of the following techniques is constrained by the TTM. Specifically, the TTM imposes the following constraints:

- The number of entries per word
- The number of memory words in the table
- 20 • The total number of entries in the table

The optimization problems discussed in this section fall into the class of *bin packing* or *knapsack* problems. See Cormen et al., *Introduction to Algorithms*, Cambridge, MA, The
 25 MIT Press, 1990. The number of entries per word defines the bin (or knapsack) size for the packing problems. The structure of the coded transition table may be altered to minimize the number of memory accesses by increasing the total number of entries and/or words required
 30 to represent the table so long as the total number entries or total number of words (bins or knapsacks) does not exceed the limits imposed by the Transition Table Memory.

Optimizations: Optimizing Throughput

35

The number of memory accesses required for a search is determined by the disposition of compressed columns in the TTM and the pattern by which those columns are accessed. The pattern depends on the set of regular expressions in the engine and the particular input data processed through the engine. In a DFA^m, m input symbols are resolved to an ECI which induces one column lookup. The number of memory accesses depends on the length of the columns in the coded transition table and the column access pattern. The column access pattern depends on the regular expression (or set of regular expressions) in the engine and the input data. The total number of memory accesses for a given search can be expressed as:

$$W = N \sum_{i=1}^{|K|} f_i w_i \quad (3)$$

where w_i is the number of words spanned by row i in Transition Table Memory, f_i is the relative frequency that row i is accessed, and N is the number of equivalence class identifiers produced by the input data.

While there is no prior knowledge of the input data, there is an ability to alter the structure of the coded transition table. By re-ordering the rows in the direct-addressed transition table, one can affect the length of the columns in the coded transition table. The optimization problem is to choose the column ordering that minimizes the total number of memory accesses, W . Assume that the transition table column access pattern follows a uniform distribution, $f_i = N/|K|$. In this case:

$$W = \frac{N^2}{|K|} \sum_{i=1}^{|K|} w_i \quad (4)$$

Under these conditions, the optimization problem is to minimize the quantity:

$$v = \sum_{i=1}^{|K|} w_i = \sum_{i=1}^{|K|} \left[\frac{\text{count}_i}{x} \right] \quad (5)$$

Recall that count_i is the transition count _{i} , or the number of entries in row i of the run-length coded transition table and x is the number of entries per word in the TTM.

To simplify the optimization problem, one can assume that
 5 $x = 1$, so the quantity that now needs to be minimized is:

$$v = \sum_{i=1}^{|K|} \text{count}_i \quad (6)$$

This will generally yield similar results to minimizing the function with an arbitrary x .

Figure 19 illustrates the benefits of state reordering
 10 for the running example presented herein.

There are many approaches to state reordering. One approach is to minimize the length of a single column of the coded transition table by ordering the rows of the direct-addressed table according to the sorted order of the entries
 15 in the row. This maximizes the efficiency of the run-length coding for that one column. However, the re-ordering may also decrease the efficiency of the run-length coding for other columns.

The preferred approach is a greedy one; preferably it is
 20 desired to maximize the length of the runs for the most columns, thereby minimizing the length of each encoded column.

One can start by creating a difference matrix, which given two states indicates the number of BCIs that differ, and so will not continue a run. This algorithm is shown in Figure
 25 20.

Next, one then orders the states from some starting point based on the entries in the difference matrix. One preferably chooses the states that preserves the most run lengths to get the next label. The starting state that is chosen is
 30 preferably the state that has the largest column-sum in the difference matrix. The idea for picking that state first is that it is the state that is the most different from all others. By moving that state to the end (rather than in the middle), one preserves the longest runs. This algorithm is
 35 outlined in Figure 21. Together the algorithms of Figures 20

and 21 serve as a "transition table state re-ordering" algorithm

Optimizations: Memory Packing

- 5 Recall that the layer of indirection allows a column of the coded transition table to begin and end at any location in the TTM. Naïve packing of coded table columns into physical memory can thwart the aforementioned optimizations by incurring an extra memory access for each table column.
- 10 Notice in Figure 15 that the run-length coded transition table column associated with input symbol '.' (ECI 1) contains three entries, but it spans two memory words in the TTM. While it is possible to store the column in a single memory word, accessing this column requires two memory accesses as laid out
- 15 in Figure 15. One can take advantage of the flexibility provided by the layer of indirection by ensuring that a coded transition table row spans at most w words, where:

$$w \leq \left\lceil \frac{\text{count}}{x} \right\rceil \quad (7)$$

- Figure 20 shows an example of using this constraint to pack
- 20 the coded transition table into the TTM. It can be seen that retrieving the column associated with ECI 1 requires only one memory access. In this example, there is no loss in memory efficiency; however this may not always be the case. For example, consider the case where a memory word holds three
- 25 entries and every coded transition table column contains two entries.

- This memory packing problem is a variant of the classical fractional knapsack problem where w is the constraint or objective function. See Black, P.E., *Dictionary of Algorithms*
- 30 *and Data Structures*, NIST, 2004, the entire disclosure of which is incorporated herein by reference. The difference in the preferred embodiment here is that we require contiguous storage of coded transition table columns. This imposes an additional constraint when partitioning an object (coded

transition table column) across multiple knapsacks (memory words) in the classical problem.

One solution to this problem is based on *subset sum*. While this is an NP-complete problem in the general case (see
 5 Garey and Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman and Co., 1979), there are certain conditions in which it runs in polynomial time, namely if the sum is much less than the number of elements that are to be chosen from to create the
 10 sum. The sum of the preferred embodiment will always be the width of a memory word, so the preferred algorithm will also run in polynomial time.

The basic idea is to find the longest run-length coded column and choose it first. One then will pack it into memory
 15 words guaranteeing that it achieves the best possible packing. One can then take the number of remaining entries in the last column and apply *subset sum* on it with the remaining run-length coded columns. This will pack the memory as full as possible without causing additional memory accesses. This
 20 process is repeated until no encoded columns remain. This algorithm is outlined in Figure 21, referred to herein as an "optimal memory word packing" algorithm, where R is the set of run-length coded columns, and w is the width of a memory word.

25 An Implementation Architecture

In this section, an implementation of a high-performance regular expression search system based on the preferred high-throughput DFA and pipelined transition table techniques is described. The focus of this implementation is a hybrid
 30 processing platform that includes multiple superscalar microprocessors and reconfigurable hardware devices with high-bandwidth interconnect to an array of high-speed disks. Figure 24 shows an example of the system-level architecture and includes a user interface, regular expression compiler,

35

file I/O controller, regular expression firmware module, and results processor.

The purpose of the architecture is to maximize throughput by embedding an array of regular expression engines in the reconfigurable hardware devices (e.g., FPGAs). The array of engines, supporting control logic, and context buffer(s) may be logically viewed as a single firmware module. In an embodiment wherein the regular expression circuits/engines are realized on an FPGA, these engines can be synthesized to a hardware definition language (HDL) representation and loaded onto the FPGA using known techniques.

Each regular expression engine's primary task is to recognize regular expressions in the input files streaming off of the high-speed disks. The set of regular expressions is preferably specified by the user through the user interface, compiled into high-throughput DFAs, and translated into a set of tables and register values by a collection of software components. The set of tables and register values are written to the firmware module prior to beginning a search. When a regular expression engine recognizes a pattern, it sends a message to the Results Processor that includes the context (portion of the file containing the pattern), starting and ending indexes of the pattern in the file, and the accepting state label. Depending on the operating environment and level of integration, the user interface may be a simple command line interface, Graphical User Interface (GUI), or language-specific API. The following subsections provide detailed descriptions of the remaining components.

30 *An Implementation Architecture: Regular Expression Compiler*

As detailed in Figure 5, the regular expression compiler receives the set of regular expressions specified by the user from the user interface. Thereafter, standard algorithms from the art are used to parse the specified regular expression(s) to create an NFA therefrom, render the NFA to a position independent DFA, and reduce the position independent DFA to a

minimal DFA. The next step performed by the regular expression compiler is to transform the conventional minimal DFA to the preferred high throughput DFA of the present invention. This step comprises the processes described above with respect to scaling the DFA to accommodate strides of m input symbols per cycle (including the determination of appropriate match and match restart flags for the transition table and the encoding of input symbols to ECIs), and run-length coding the transition table. Next, the algorithms of Figures 20 and 21 can be run to optimize the storage of the run-length coded transitions in transition table memory and the algorithm of Figure 23 can be run to optimize the packing of the run-length coded transition entries into words of the transition table memory. Thereafter, the entries in the indirection table can be determined, and the compiler is ready to issue commands to the regular expression circuit 502 that will operate to populate the circuit's memory for the input symbol-to-ECI tables, the indirection memory table and transition table.

20

An Implementation Architecture: Results Processor

It is expected that any of a variety of techniques can be used to report the results of a search via a results processor. The preferred results processor can be configured to resolve the exact expression and input string segment for each match using the results produced by the regular expression circuits (engines). In a preferred embodiment such as that shown in Figure 24, the results produced by a regular expression circuit (engine) include a unique engine identifier (ID), the start state, the accepting state, and the ECI for m input symbols that triggered the accepting transition. The results may also include the context for the match, which is a section of the input stream containing a string that matches the pattern. The results processor reports the matching string and associated regular expression (pattern) to the user interface.

An Implementation Architecture: File I/O Controller

The file I/O controller is a component of the system that controls the input stream. In the exemplary system of Figure 24, the file I/O controller controls the stream of files flowing from a data store to the regular expression circuits. Note that the input stream may also be fed by a network interface (or other data interface) as is known in the art.

10 *An Implementation Architecture: Regular Expression Firmware*

The regular expression firmware module is the primary datapath component in the system architecture shown in Figure 24. It preferably contains an array of regular expression engines (or pattern matching circuits) and a small amount of control logic. The number of engines in the array depends on the capacity of the reconfigurable hardware devices in the system. The control logic broadcasts the input file stream to each regular expression engine, thus the engines operate in parallel on the same input data. The control logic also sends a copy of the input to a context buffer. The size of the context buffer depends on the amount of context that is to be sent to the Results Processor when an engine recognizes a pattern. This parameter may be tuned to maximize the amount of context that may be returned while not overloading the firmware/software interface.

As previously mentioned, the throughput of a regular expression engine is fundamentally limited by the rate at which it can compute state transitions for the deterministic finite automaton. Resolving the next state based on the current state and input symbol is an inherently serial operation. In order to take advantage of the reconfigurable logic resources available on the preferred implementation platform, it is desired to maximize parallelism. Pipelining is a common technique for increasing the number of parallel operations in serial computations; however, it requires that the processing pipeline be free of feedback loops. The

outputs of operations at a given stage of the pipeline cannot depend upon the results of a stage later in the pipeline. As shown in Figure 25, the regular expression engine is a series of pipelined computational blocks. Note that there are no
 5 feedback loops between the blocks; each block operates on the results of the previous block only. This is a distinguishing feature of the preferred embodiment of the present invention. Alphabet encoding is a state-independent operation that only operates on the set of input symbols. Indirection table
 10 lookups use the resulting input symbol ECI to locate an entry. Transition table lookups depend only on the pointer and indexes contained in the Indirection Table entry. The only operation that depends on the current state is the last computation in the State Select block. By effectively
 15 "pushing" this singular feedback loop to the final stage of the pipeline, the preferred embodiment maximizes parallelism and hence the throughput of the regular expression engine. The following sub-sections describe the design of each block in the regular expression engine.

20

Regular Expression Firmware: Alphabet Encoding

The Alphabet Encoding block assigns an Equivalence Class Identifier (ECI) for a set of m input symbols. If each input symbol is specified using i bits and an ECI is specified using
 25 p bits, then the Alphabet Encoding block essentially reduces the input from mi bits to p bits. A straightforward method for performing this operation is to perform pairwise combinations using direct-addressed tables. As shown in Figure 26, the first set of tables transforms one i bit input
 30 symbol to a j bit ECI. This step maps single input symbols to equivalence classes. The next stage of tables generates a k bit ECI for two input symbols by simply concatenating two j bit ECIs for single symbols and direct-addressing an ECI table. Note that j is upper bounded by the addressability of
 35 the memory used to implement the ECI tables in the second stage. Specifically, $2j$ must be less than or equal to the

number of address bits supported by the memory. Similarly, the next stage generates a p bit ECI for four input symbols by concatenating two k bit ECIs for two symbols and direct-addressing an ECI table. Likewise, the address space supported by the ECI table places an upper bound on k . In theory, this technique may be used to assign an ECI to any number of input symbols; but in practice, the memory efficiency significantly degrades as the number of symbols covered by the final equivalence class increases.

The logic required to implement subsequent stages of the Figure 25 pipeline have already been discussed above.

Regular Expression Firmware: Buffers

Each regular expression engine preferably includes small input and output buffers. The input buffers prevent a single engine from stalling every engine in the array when it must retrieve a transition table column that spans multiple words in the TTM. While the entire array must stall when any engine's input buffer fills, the input buffers help isolate the instantaneous fluctuations in file input rates. The output buffers allow the regular expression engines to continue processing after it has found a match and prior to the match being transmitted to the Results Processor. The Context Buffer preferably services the output buffers of the regular expression engines in round-robin fashion. If the output buffer of any engine fills, then the engine must stall prior to sending another result to the output buffer. The array preferably must stall if the engine's input buffer fills.

While the present invention has been described above in relation to its preferred embodiment, various modifications may be made thereto that still fall within the invention's scope. Such modifications to the invention will be recognizable upon review of the teachings herein. For example, while the transition tables have been described herein such that the rows correspond to states and the columns

correspond to ECIs, it should be readily understood that the rows and columns of any of the tables described herein can be reversed. As such, the full scope of the present invention is to be defined solely by the appended claims and their legal
5 equivalents.

CLAIMS:

1. In a device for matching an input string to a pattern via a deterministic finite automaton (DFA), the DFA comprising a plurality of states including a current state and a plurality of possible next states, the input string comprising a plurality of input symbols, the improvement comprising:

the device comprising at least two parallel pipeline stages;

a first one of the pipeline stages being configured to retrieve a plurality of transitions to a possible next state of the DFA from a pre-populated memory; and

a second one of the pipeline stages that is configured to choose, based at least in part upon the DFA's current state, one of said retrieved transitions from which the integrated circuit will determine the next state of the DFA, wherein the second one of the pipeline stages is downstream from the first one of the pipeline stages.

2. The device of claim 1 wherein the improvement further comprises the first one of the pipeline stages being configured to retrieve a plurality of transitions to a possible next state of the DFA from a pre-populated memory without consideration of the current state of the DFA.

3. The device of claim 2 wherein the improvement further comprises the second one of the pipeline stages being the final downstream pipeline stage of the device.

4. A method for matching an input string to a pattern in a device that realizes a deterministic finite automaton (DFA), the DFA comprising a plurality of states including a current state and a plurality of possible next states, the input string comprising a plurality of input symbols, the device comprising at least two parallel pipeline stages, the method comprising:

within one of the pipeline stages, retrieving from a memory a plurality of stored transitions to a next possible state of the DFA; and

within another of the pipeline stages that is downstream in the pipeline from the pipeline stage from which the plurality of transitions were retrieved,

selecting one of the retrieved transitions to identify the next state of the DFA, wherein the selecting is based at least in part upon the current state of the DFA.

5. The method of claim 4 wherein the retrieving step is performed without consideration of the current state of the DFA.

6. A device comprising:

a multi-stage pattern matching pipeline configured to receive an input data stream comprising a plurality of input symbols and process the received data stream, wherein the pipeline realizes a pattern matching deterministic finite automaton (DFA) that is configured to determine whether any string of input symbols within the data stream matches a pattern, the DFA having a plurality of states including a current state, the pipeline comprising a plurality of stages including a final stage, each stage having at least one stage input and at least one stage output, the final stage being configured to provide as outputs a next state for the DFA, wherein the only stage of the pipeline that receives a feedback loop based on a stage output is the final stage.

7. The device of claim 6 wherein the final stage is configured to receive the next state in the feedback loop such that the next state is used as the current state when processing a next input symbol of the input data stream.

8. The device of claim 7 wherein the pipeline is further configured with a plurality of the multi-stage pattern matching pipelines for processing the input data stream in parallel.

9. The device of claim 7 wherein the pipeline comprises an FPGA.

10. The device of claim 7 wherein the pipeline comprises an ASIC.

11. The device of claim 7 wherein the pipeline comprises a processing system, the processing system comprising a plurality of different processors that are

arranged in a pipeline, each processor being configured to implement at least one of the pipeline stages.

12. A device comprising:

a data processing pipeline configured to receive an input data stream comprising a plurality of input symbols and process the received data stream through a logic circuit to determine whether any input symbol string within the input data stream matches a pattern, the logic circuit being configured to realize a deterministic finite automaton (DFA), the DFA comprising a plurality of states, an alphabet of input symbols, an alphabet of equivalence class identifiers (ECIs), a set of functions that map groups of m input symbols to corresponding ECIs, and a set of functions that define transitions between states of the DFA, each state transition function being indexed by an ECI and comprising a next state identifier and a match flag, wherein m is an integer that is greater than or equal to one.

13. The device of claim 12 wherein the DFA further comprises a set of functions that map each ECI to at least one state transition function.

14. The device of claim 13 wherein each state transition function further comprises a match restart flag.

15. The device of claim 14 wherein the DFA further comprises a variable that tracks the starting input symbol of each potential pattern match.

16. The device of claim 15 wherein the DFA further comprises a variable that tracks the final input symbol of an actual pattern match.

$$\backslash \$[0-9] + (\backslash .[0-9]\{2\})\{0,1\}$$

Figure 1 – PRIOR ART

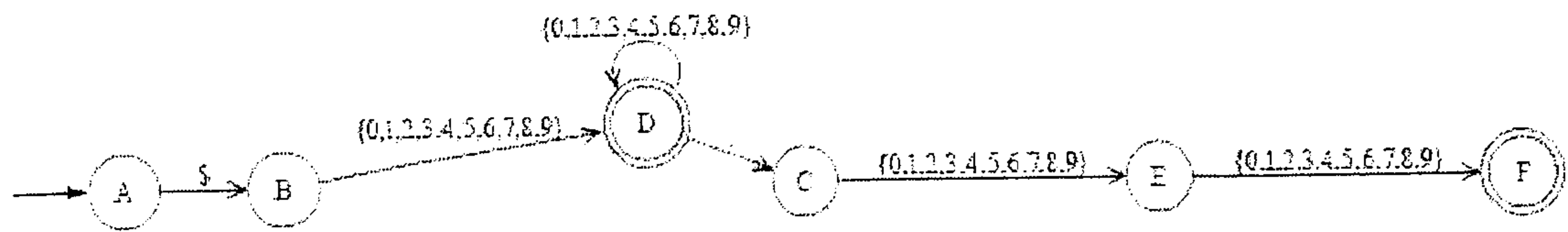


Figure 2 – PRIOR ART

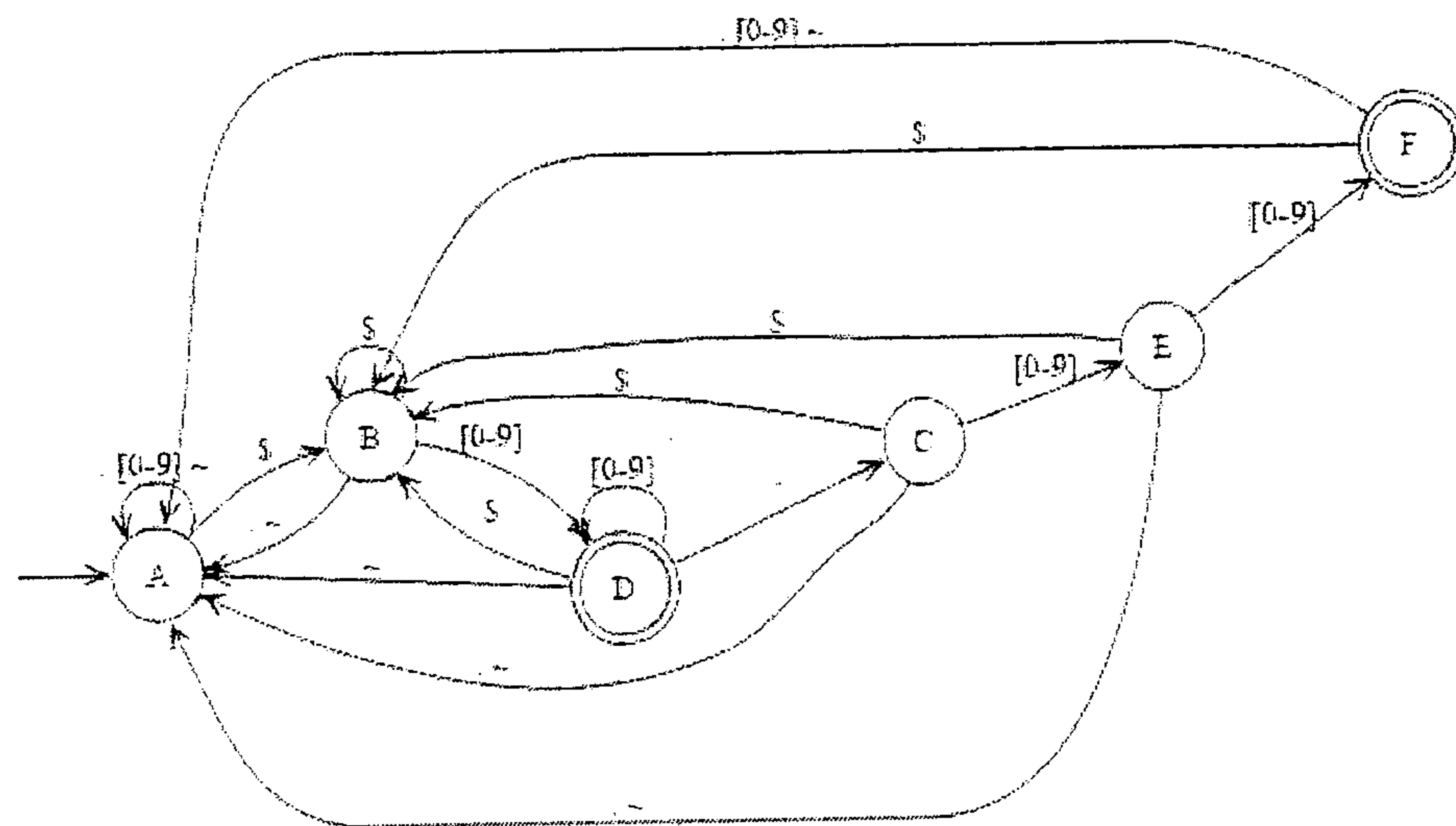


Figure 3 – PRIOR ART

State	Symbol			
	{0...9}	.	\$	~...
A	A...A	A	B	A...
B	D...D	A	B	A...
C	E...E	A	B	A...
D	D...D	C	B	A...
E	F...F	A	B	A...
F	A...A	A	B	A...

Figure 4 – PRIOR ART

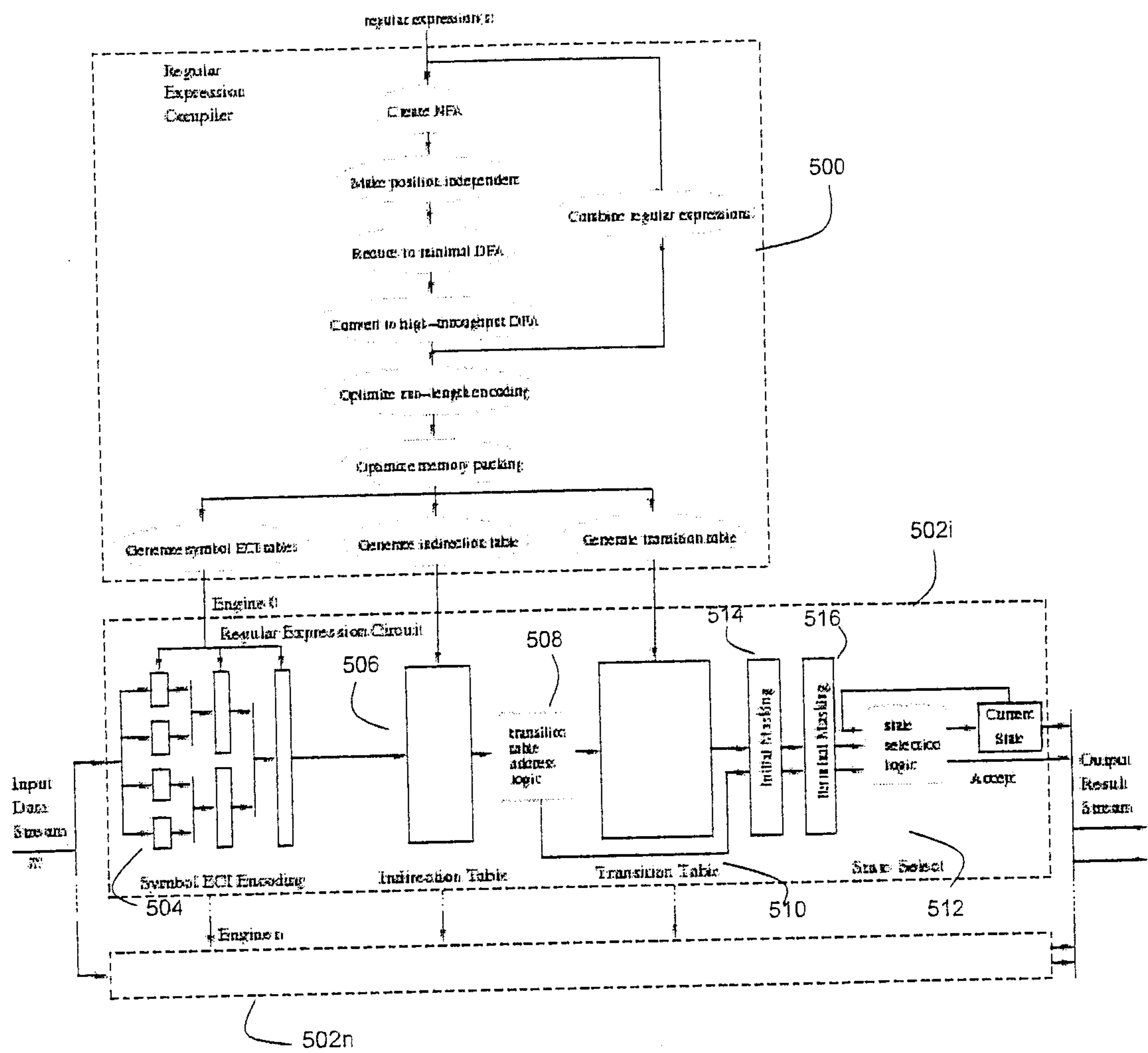


Figure 5

3/11

```

function DOUBLESTRIDE( $\delta^k$ ):  $\delta^{2k}$ 
  foreach  $q \in Q$  do
    foreach  $a \in \Sigma^k$  do
      foreach  $b \in \Sigma^k$  do
         $\delta^{2k}(q, ab) \leftarrow \delta^k(\delta^k(q, a), b)$ 
      end
    end
  end

```

Figure 6

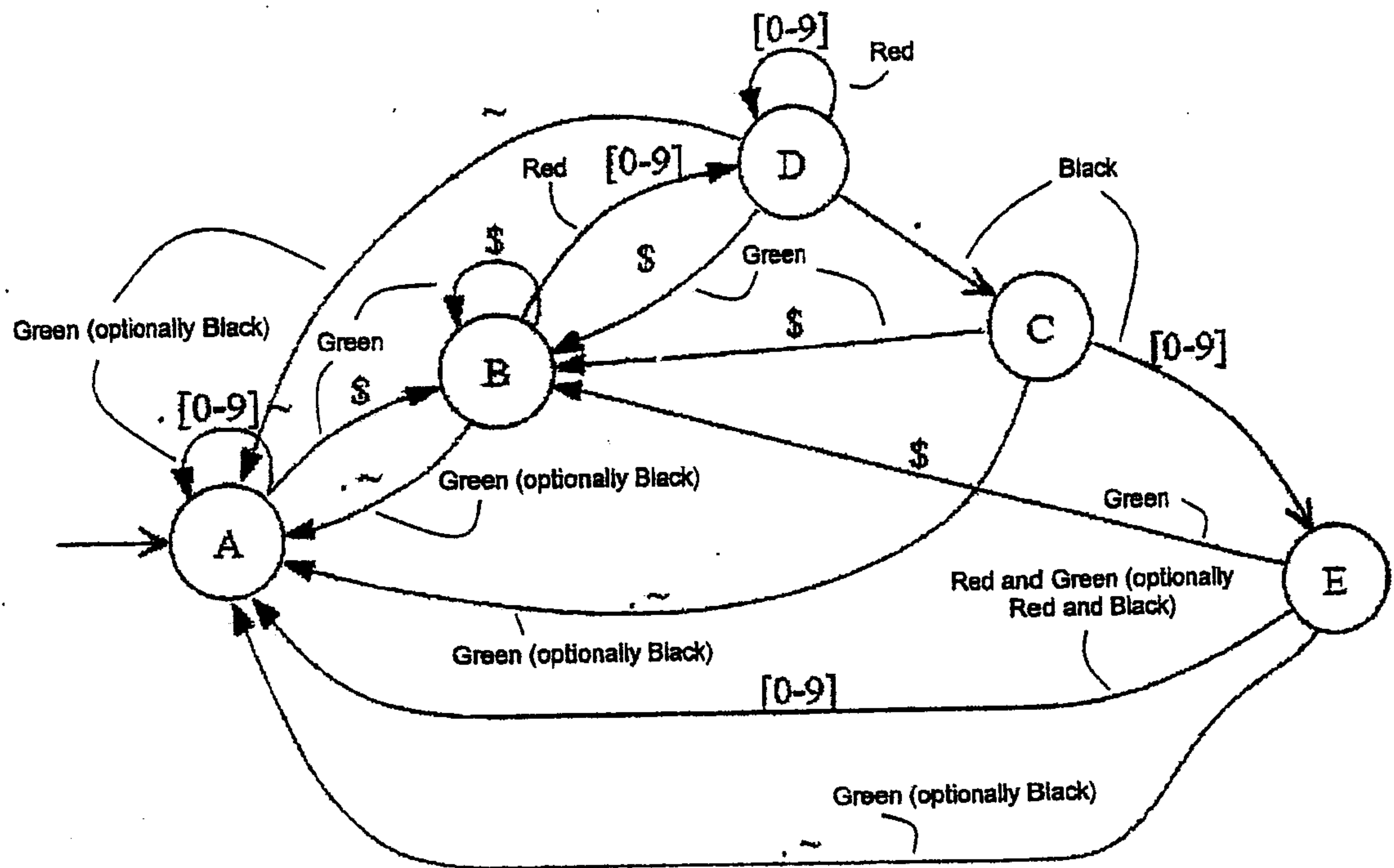


Figure 7

4/11

```

function PARTITION( $Q, \Sigma, \delta : Q \times \Sigma \mapsto Q$ ): ( $result, K, \kappa : \Sigma \mapsto K$ )
   $WorkList \leftarrow \Sigma$ 
   $result \leftarrow \emptyset$ 
   $i \leftarrow -1$ 
  while  $WorkList \neq \emptyset$  do
     $w \leftarrow$  some element of  $WorkList$ 
     $equiv \leftarrow \{a \in \Sigma \mid \forall q \in Q \delta(q, a) = \delta(q, w)\}$ 
     $i \leftarrow i + 1$ 
     $K \leftarrow K \cup \{i\}$ 
    foreach  $sym \in equiv$  do  $\kappa(sym) \leftarrow i$ 
     $result \leftarrow result \cup equiv$ 
     $WorkList \leftarrow WorkList - equiv$ 
  end

```

Figure 8

Symbol	ECI	State	ECI			
			0	1	2	3
\$	0	A	(B,1,0)	(A,1,0)	(A,1,0)	(A,1,0)
.	1	B	(B,1,0)	(A,1,0)	(D,0,1)	(A,1,0)
0-9	2	C	(B,1,0)	(A,1,0)	(E,0,0)	(A,1,0)
all else	3	D	(B,1,0)	(C,0,0)	(D,0,1)	(A,1,0)
		E	(B,1,0)	(A,1,0)	(A,0,1)	(A,1,0)

(a) (b)

Figure 9

5/11

Symbols	ECI
0	0
0	1
0	2
0	3
1	0
1	1
1	2
1	3
2	0
2	1
2	2
2	3
3	0
3	1
3	2
3	3

(a)

State	ECI							
	0	1	2	3	4	5	6	7
A	(B,1,0)	(A,1,0)	(D,1,1)	(A,1,0)	(B,1,0)	(A,1,0)	(A,1,0)	(A,1,0)
B	(B,1,0)	(A,1,0)	(D,1,1)	(A,1,0)	(B,1,1)	(C,0,1)	(D,0,1)	(A,1,1)
C	(B,1,0)	(A,1,0)	(D,1,1)	(A,1,0)	(B,1,0)	(A,1,0)	(A,0,1)	(A,1,0)
D	(B,1,0)	(A,1,0)	(D,1,1)	(E,0,0)	(B,1,1)	(C,0,1)	(D,0,1)	(A,1,1)
E	(B,1,0)	(A,1,0)	(D,1,1)	(A,1,0)	(B,1,1)	(A,1,1)	(A,1,1)	(A,1,1)

(b)

Figure 10

```

function BASEDFA( $Q, \Sigma, q_0, \delta, A, \rho$ ) :  $DFA^1 = (Q', \Sigma', q'_0, K, \kappa, \delta')$ 
   $K \leftarrow \{0\}$ 
  foreach  $a \in \Sigma$  do  $\kappa(a) \leftarrow 0$ 
   $WorkList \leftarrow \{q_0\}$ 
  while  $WorkList \neq \emptyset$  do
     $w \leftarrow$  some element of  $WorkList$ 
     $WorkList \leftarrow WorkList - \{w\}$ 
    /*  $w$  is a set of states that is already  $\lambda$ -closed
    /*  $q_0 \in w$  by construction
     $Q' \leftarrow Q' \cup \{w\}$ 
    foreach  $a \in \Sigma$  do
       $targs \leftarrow \{t \mid (\exists s \in w) \delta(s, a) = t\}$ 
       $targs \leftarrow \text{LAMBDA CLOSURE}(targs)$ 
       $red \leftarrow targs \cap A \neq \emptyset$ 
       $green \leftarrow (\forall s \in w - \{q_0\}) (\forall t \in Q - \{q_0\}) \delta(s, a) \neq t$ 
       $\delta' \leftarrow \delta' \cup \{w \times a \mapsto (targs, green, red)\}$ 
      if  $targs \notin Q'$ 
      then  $WorkList \leftarrow WorkList \cup \{targs\}$ 
  end

```

Figure 11

6/11

State	ECI			
	0	1	2	3
A	5(B,1,0)	3(A,1,0)	1(A,1,0)	5(A,1,0)
B			1(D,0,1)	
C			1(E,0,0)	
D		1(C,0,0)	1(D,0,1)	
E		1(A,1,0)	1(A,0,1)	

Figure 12

5	1	1	0	3	0	1	0	1	2	0	0	1	0	1	0	1	0	1	3	0	1	1	4	0	0	1	3	0	1	1	0	0	1	5	0	1	0
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Figure 13

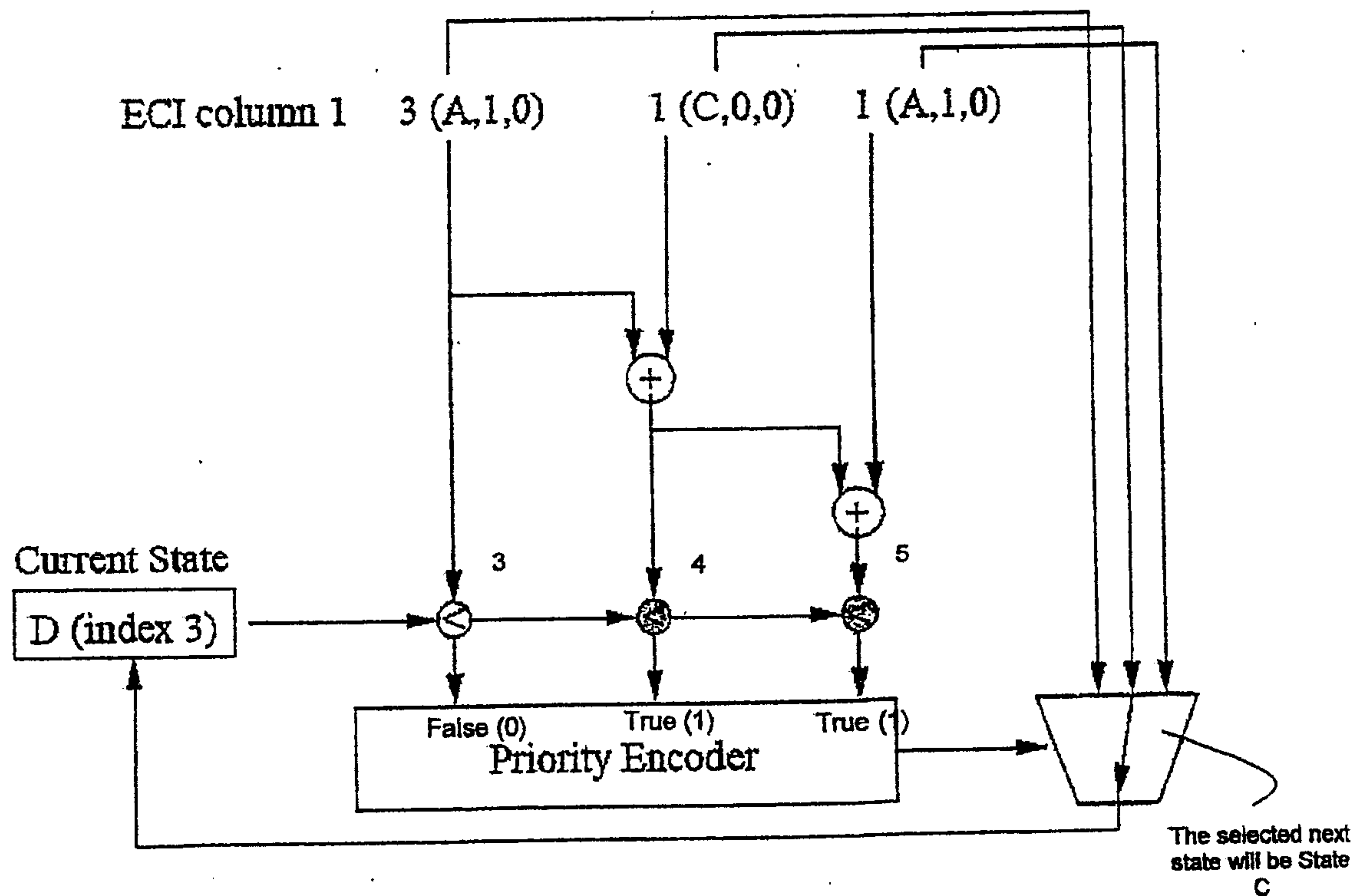


Figure 14

7/11

ECI	pointer	index	count
0	0	0	1
1	0	1	3
2	1	1	5
3	3	0	1

Indirection Table

(a)

addr			
0	5 (B,1,0)	3 (A,1,0)	1 (C,0,0)
1	1 (A,1,0)	1 (A,1,0)	1 (D,0,1)
2	1 (E,0,0)	1 (D,0,1)	1 (A,0,1)
3	5 (A,1,0)		

Transition Table Memory

(b)

Figure 15

ECI	pointer	index	count
0	0	0	1
1	0	1	3
2	1	1	5
3	3	0	1

Indirection Table

(a)

addr			
0	5 (B,1,0)	3 (A,1,0)	1 (C,0,0)
1	5 (A,1,0)	1 (A,1,0)	2 (D,0,1)
2	3 (E,0,0)	4 (D,0,1)	5 (A,0,1)
3	5 (A,1,0)		

Transition Table Memory

(b)

Figure 16

symbol	ECI	pointer	initial index	terminal index	word count
\$	0	0	0	0	0
.	1	0	1	0	1
[0-9]	2	1	1	2	1
~	3	3	0	0	0

Indirection Table

addr			
0	5 (B,1,0)	3 (A,1,0)	1 (C,0,0)
1	5 (A,1,0)	1 (A,1,0)	2 (D,0,1)
2	3 (E,0,0)	4 (D,0,1)	5 (A,0,1)
3	5 (A,1,0)		

Transition Table Memory

Figure 17

8/11

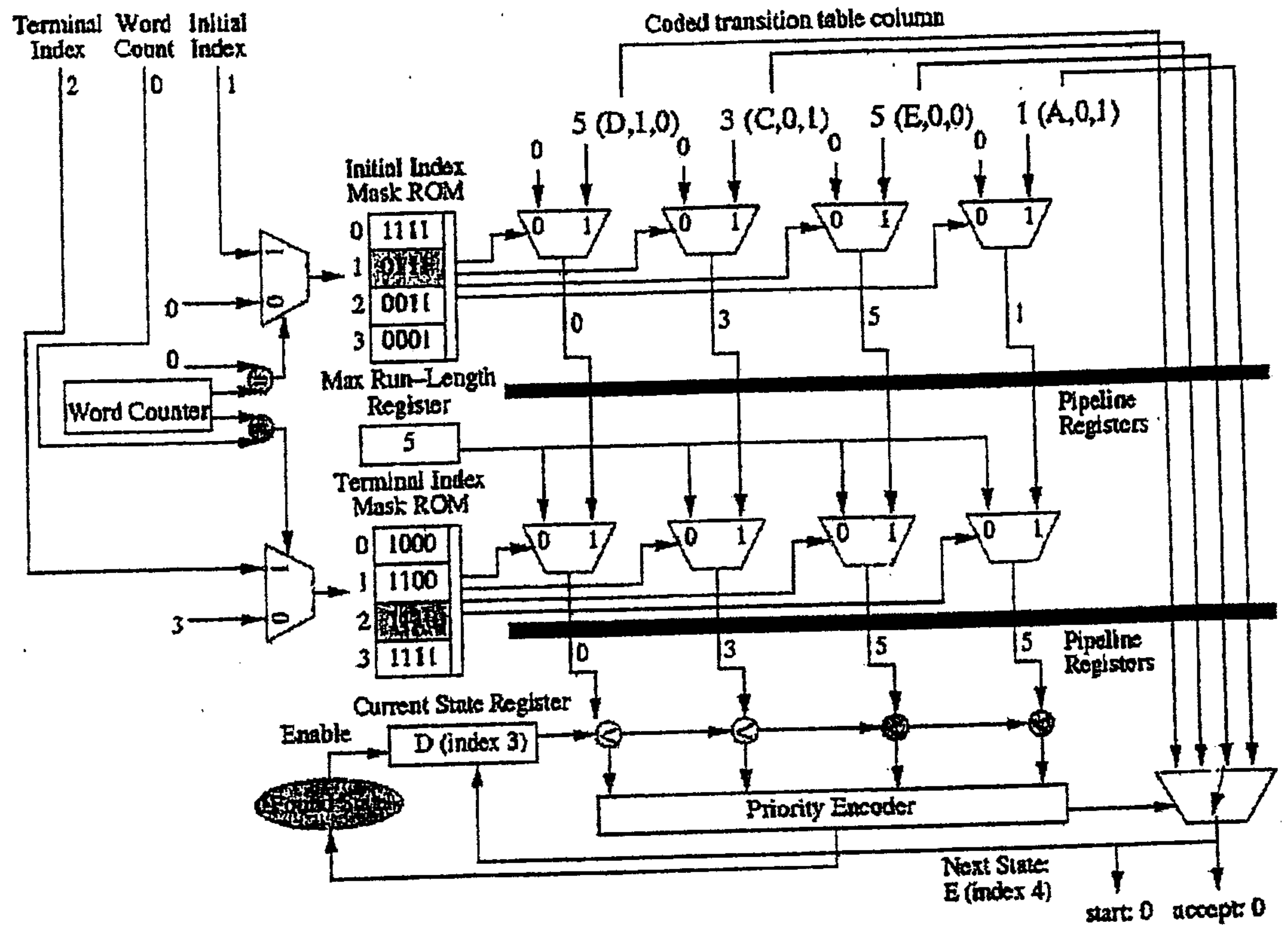


Figure 18

State	ECI				State	BCI			
	0	1	2	3		0	1	2	3
A	(B,1,0)	(A,1,0)	(A,1,0)	(A,1,0)	A	(5,B,1)0	(4,A,1)0	(1,A,1)0	(5,A,1)0
C	(B,1,0)	(A,1,0)	(E,0,0)	(A,1,0)	C			(1,E,0)0	
E	(B,1,0)	(A,1,0)	(A,0,1)	(A,1,0)	E			(1,A,0)1	
B	(B,1,0)	(A,1,0)	(D,0,1)	(A,1,0)	B			(2,D,0)1	
D	(B,1,0)	(C,0,0)	(D,0,1)	(A,1,0)	D		(1,C,0)0		

(a)

(b)

Figure 19


```
function DIFFMATRIX( $Q, \Sigma, \delta$ ):  $D$ 
  foreach  $q \in Q$  do
    foreach  $p \in Q$  do
      foreach  $s \in \Sigma$  do  $D[q][p] \leftarrow D[q][p] + \delta(q, s) \neq \delta(p, s)$ 
    end
  end
```

Figure 20

```
function RLOPTIM( $Q, q_0, \Sigma, \delta, D$ ):  $DFA' = (Q', \Sigma, q'_0, \delta')$ 
   $Used \leftarrow \emptyset$ 
   $cur \leftarrow \text{MAX}(\sum_{i \in Q} D[i])$ 
  while  $Used \neq Q$  do
     $next \leftarrow \text{MAX}(D[cur] - Used)$ 
     $Used \leftarrow Used \cup \{next\}$ 
     $cur \leftarrow next$ 
  end
```

Figure 21

symbol	ECI	pointer	initial index	terminal index	word count	addr
\$	0	0	0	0	0	0
.	1					1
[0-9]	2	1	0	0	1	2
~	3	2	1	1	0	3

Indirection Table

5 (B.1,0)		
1 (A,1,0)	1 (E,0,0)	1 (A,0,1)
2 (D,0,1)	5 (A,1,0)	

Transition Table Memory

Figure 22

10/11

```

procedure COLPACK( $R, x$ )
  while  $R \neq \emptyset$  do
     $longest \leftarrow \text{MAXLENGTH}(R)$ 
     $R \leftarrow R - \{longest\}$ 
    PACK( $longest$ )
     $sum \leftarrow x - (longest \bmod x)$ 
     $L_0 \leftarrow \{0\}$ 
    foreach  $col \in R$  do  $L_i \leftarrow \text{MERGELISTS}(L_{i-1}, L_{i-1} + \text{LENGTH}(col))$ 
    remove from  $L_i$  every element that is greater than  $sum$ 
     $S \leftarrow \text{Max}(L_n)$ 
    foreach  $col \in S$  do
       $R \leftarrow R - \{col\}$ 
      call PACK( $col$ )
  end

```

Figure 23

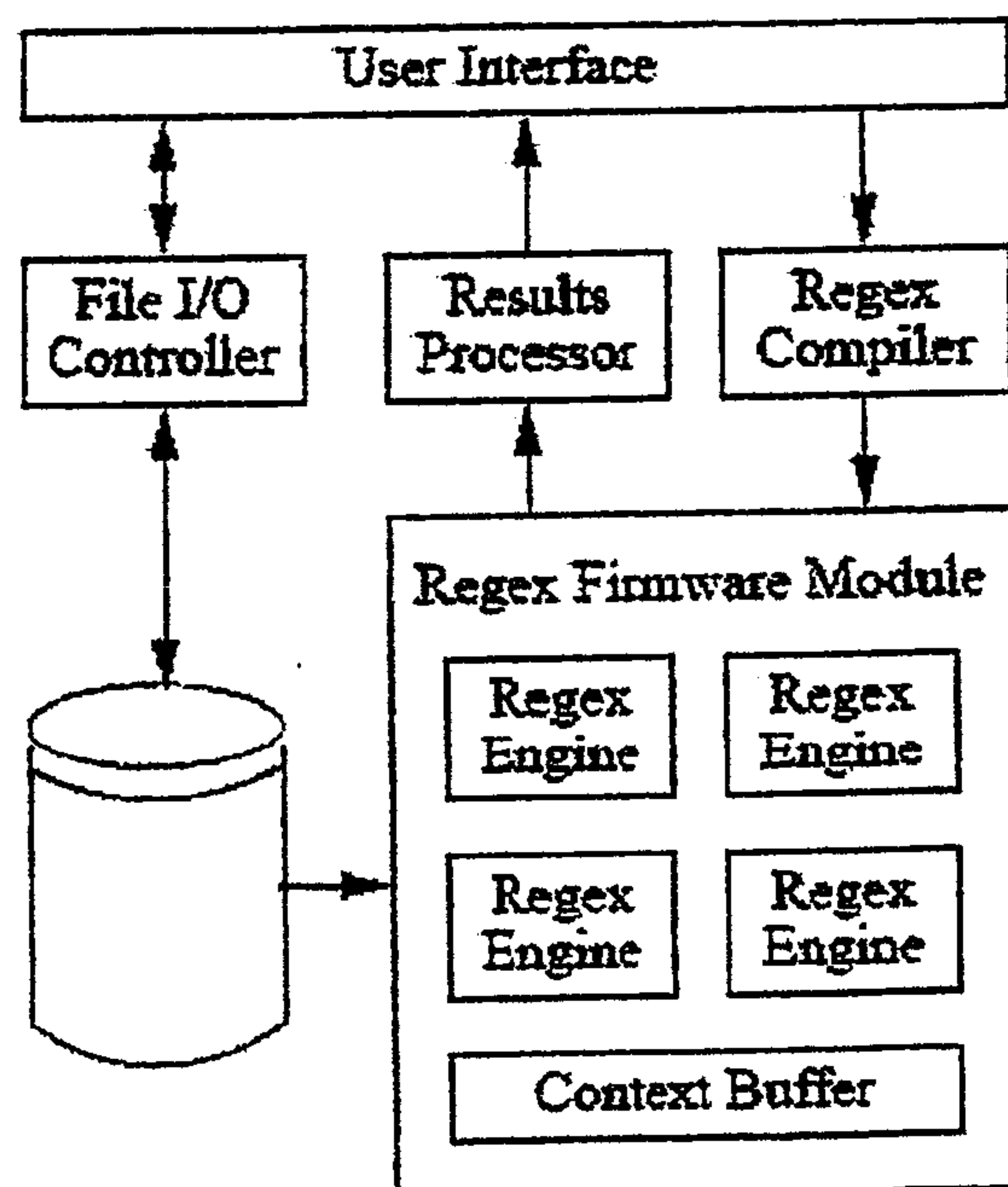


Figure 24

11/11

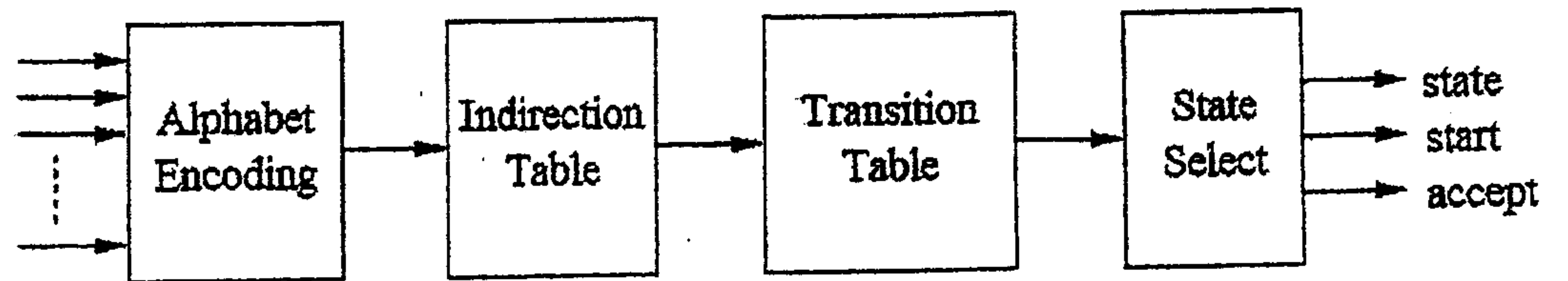


Figure 25

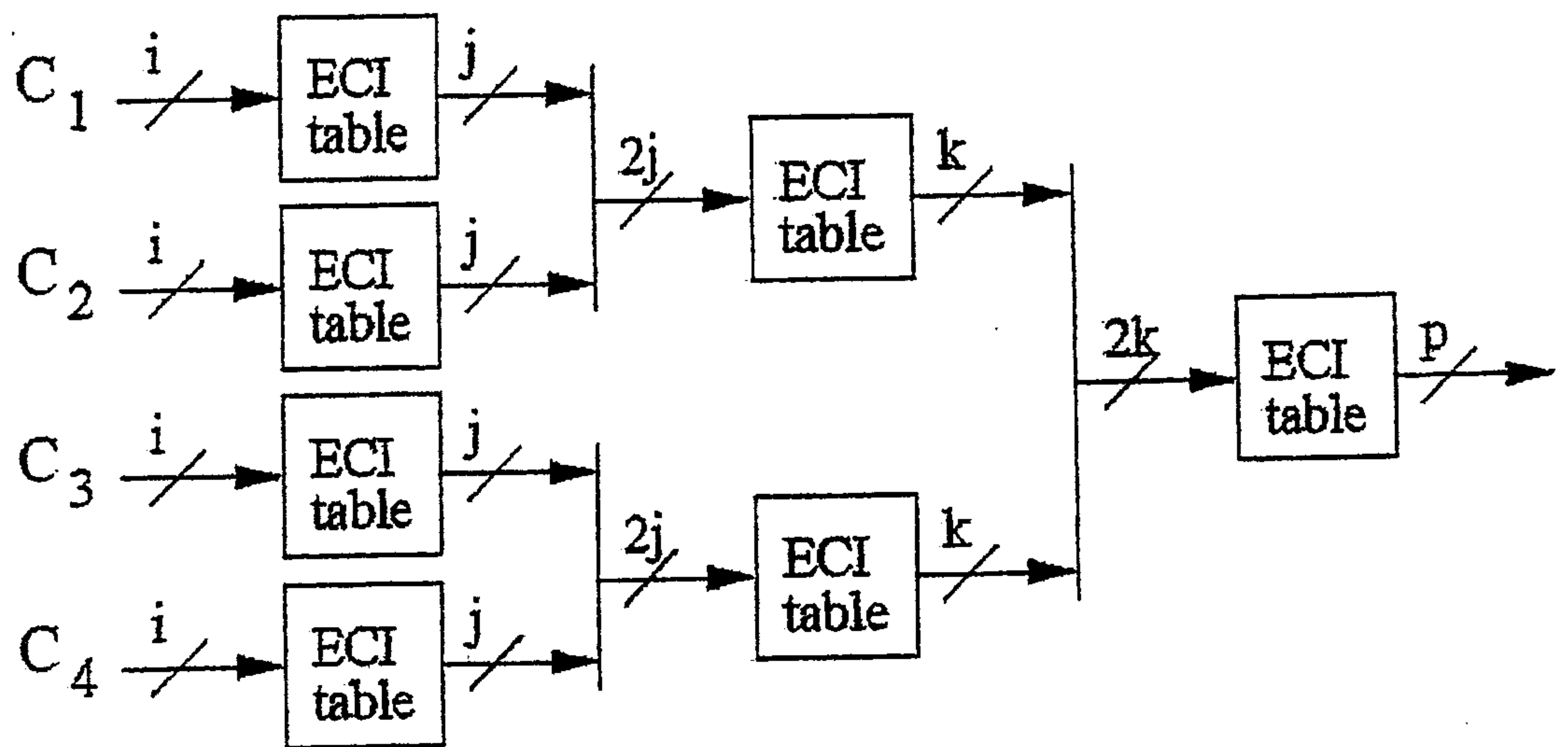


Figure 26