EP 0041549.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 3:
G06F 3/00

A1

(11) International Publication Number: WO 81/01758

(43) International Publication Date: 25 June 1981 (25.06.81)

(21) International Application Number: PCT/US80/01597

(22) International Filing Date: 26 November 1980 (26.11.80)

(31) Priority Application Number: 101,319

(32) Priority Date: 7 December 1979 (07.12.79)

(33) Priority Country: US

(71) Applicant: NCR CORPORATION [US/US]; World Headquarters, Dayton, OH 45479 (US).

(72) Inventors: OOSTERBAAN, DuWayne, Dale; 1351 Felicita Lane, Escondido, CA 92025 (US). WILLIAMS, Gerard, John; 26984 Banbury Drive, Valley Center, CA 92082 (US).

(74) Agents: DUGAS, Edward, et al; Patent Division, NCR Corporation, World Headquarters, Dayton, OH 45479 (US).

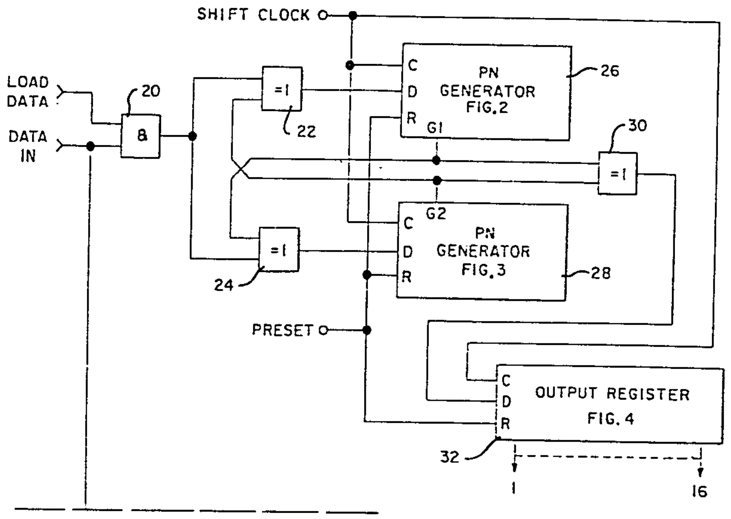
(81) Designated States: CH (European patent), DE (European patent), FR (European patent), GB (European patent), JP.

Published
With international search report

(54) Title: APPARATUS AND METHOD FOR HASHING KEY DATA

(57) Abstract

Hashing of a key data signal is accomplished by utilizing a pseudo-random number signal generator (22, 24, 26, 28 and 30) for generating a randomized signal in response to shift signals and the key data signals and an output register (32) for serially receiving the generated pseudo-ramdom signal and for providing segments of the serially-received signal at its output. A counting circuit (36, 56) responsive to a preselected number of shift signals provides an output valid signal when the preselected number of shift signals has occurred and further shifts the pseudo-random number signal generator an amount corresponding to the preselected number of shift signals. The pseudo-random number signal generator includes a pair of cross-coupled shift registers (26, 28). The method of hashing the key data utilizes the steps of presetting the pseudo-ramdon number generator and the counting circuit to an initialized state. The counting circuit is then loaded with a predetermined count whereupon key data is entered into the pseudo-random number generator so as to randomize the key data. A valid signal is provided when a block of key data has been hashed and the steps of entering the key data and providing a valid signal upon the occurrence of each block of key data is repeated until all key data blocks have been hashed.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria
AU	Australia
BR	Brazil
CF	Central African Republic
CG	Congo
CH	Switzerland
CM	Cameroon
DE	Germany, Federal Republic of
DK	Denmark
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
HU	Hungary
JP	Japan
KP	Democratic People's Republic of Korea
LI	Liechtenstein
LU	Luxembourg
MC	Monaco
MG	Madagascar
MW	Malaŵi
NL	Netherlands
NO	Norway
RO	Romania
SE	Sweden
SN	Senegal
SU	Soviet Union
TD	Chad
TG	Togo
US	United States of America

APPARATUS AND METHOD FOR HASHING KEY DATA

Technical Field

This invention relates to apparatus for hashing key data. The invention also relates to a method for hashing key data.

Background Art

In computerized processing of data, it is common practice to store like data items as multiple entries within a named data file. A portion of each record, referred to as the key, is used to reference a specific record. The keys are assumed to be unique throughout the file. Fundamental to the processing of the data file is the search for a data record associated with a specific key. A number of techniques have been developed which perform this particular function. A class of these techniques is referred to as hashing access methods.

A hashing access method is commonly used when the number of actual keys is a small percentage of the total number of possible keys. This generally occurs when the key data is represented as ASCII character codes. An example is a 6-digit part number ranging from 000000 to 999999, which requires a 6-byte field (48 bits) with only ten valid values for each byte out of a possible 256 unique values. Another example is the use of a person's name as the key. In this case a fixed length field (say 20 bytes) is allocated for key data. Since all names do not contain 20 characters and certain combinations of letters do not realistically represent a name, a high percentage of possible bit configurations will never be used as valid keys.

A distinguishing property of hashing methods is that they do not uniquely map keys to record storage locations. Instead, they provide for more than one key to map to a specific table entry which contains the location of one or more records. The object of effective hashing methods is to arrive at a uniform distribution of the number of keys which map to a specific starting pointer thus minimizing the search time for any randomly selected key.

Hashing techniques are known from the publication by D. E. Knuth "The Art of Computer Programming", Volume 3, pages 506-549 (Addison-Wesley Publishing Company, 1973). The techniques disclosed in said publication are software methods based on algorithms. Such methods are tailored to specific sets of properties possessed by the keys, that is, alpha keys, alphanumeric keys, numeric keys, closeness of adjacent keys, number of repeated characters in the keys, etc. Thus these known hashing techniques have the disadvantage that different hashing methods are needed to effectively hash different types of key data.

Disclosure of the Invention

It is an object of the present invention to provide apparatus and a method for hashing key data which is of general application.

It is a further object of the invention to provide a hardware implementation for hashing key data.

Therefore, according to the present invention there is provided apparatus for hashing key data, characterized by pseudo-random bit generating means responsive to key data signals applied thereto to generate a randomized output signal, output register means adapted to receive said randomized output signal and control means adapted to provide an output valid signal indicative of valid hashed key data being available at an output of said output register means.

According to another aspect of the present invention, there is provided a method for hashing key data, characterized by the steps of: (a) presetting pseudo-random bit generating means and a counting device to an initialized state; (b) setting a predetermined load count into said counting device; (c) applying key data to said pseudo-random bit generating means; and (d) generating a randomized output signal by applying shift signals to said pseudo-random bit generating means in accordance with the loaded count.

An advantage of the present invention is that the hashed key data produced is such that all the original properties of closeness, adjacency and orderliness are removed, regardless of the nature of the original key data.

A further advantage is that the hashed key data has a high degree of complexity, that is, a large number of permutations have to be tried before the hashed data can be decoded. Thus a high degree of security against unauthorized access to data is provided.

Brief Description of the Drawings

One embodiment of the present invention will now be described by way of example with reference to the accompanying drawings, wherein like characters indicate like parts, and in which:

Fig. 1 is a block schematic of the preferred apparatus of the present invention.

Fig. 2 is a logic schematic of a first generator which may be used in the preferred apparatus of Fig. 1.

Fig. 3 is a logic schematic of a second generator which may be used in conjunction with the first generator of Fig. 2.

Fig. 4 is a logic schematic of a register which may be used in the preferred apparatus of Fig. 1.

Fig. 5 is a timing diagram useful in understanding the operation of the preferred apparatus embodiment of the present invention.

Mode for Carrying Out the Invention

Referring to Fig. 1; the AND gate 20 receives two inputs; a LOAD DATA input and a DATA IN input. The output of AND gate 20 is directed to inputs of Exclusive OR gates 22 and 24. The outputs of Exclusive OR gates 22 and 24 are directed to the inputs of pseudo-random number generators 26 and 28, respectively. The generator inputs are labeled D. Each generator has a clocking input, labeled C for receiving a SHIFT CLOCK signal, and a preset input, labeled R, for receiving a PRESET signal. A clocking signal applied to input C causes the generators, which are in the preferred embodiment shift registers, to shift the contents through the register. The PRESET signal initializes the generators to a starting condition. The output of generator 26 is directed to an input of an Exclusive OR gate 30 and to the input of Exclusive OR gate 24. The output from generator 28 is directed to the other input of Exclusive OR gate 30 and to an input of Exclusive OR gate 22. The output of Exclusive OR gate 30 is directed to the D input of an output register 32. The clocking input C of output register 32 is adapted to receive the SHIFT CLOCK signal. The output register 32 in the preferred embodiment of the invention is a serial-in, parallel-out shift register of sixteen stages with each of the output stages labeled from 1-16. A more detailed description of cross-coupled pseudo-random number generators is contained in international (PCT) publication No. W080102349 in the name of the present Applicant.

A serial-in, parallel-out count register 36 is adapted to receive the DATA IN signal on its D input. A LOAD COUNT signal is addressed to an input of an AND gate 34 and gated by a CLOCK signal applied to the other input of AND gate 34 to the clock input, labeled C of count register 36. A Down counter 38 receives the PRESET signal on its PE labeled input and a COUNT signal on its C labeled input. The PRESET signal is inverted by an inverter 40 to provide a $\overline{\text{PRESET}}$ signal and the CLOCK signal is inverted by an inverter 42 to provide a

$\overline{\text{CLOCK}}$ signal. The output of the down counter 38 which output is the terminal count TC is directed to an inverter 44. The output of inverter 44 is connected to an input of AND gate 46. The output of AND gate 46 is connected to the J and K inputs of a JK flip-flop 50. The LOAD DATA signal is directed to an inverting amplifier 48, the output of which is connected to the set input, labeled S, of flip-flop 50, and to an input of an AND gate 54. The Q output of flip-flop 50 is connected as an input to AND gate 52 and to AND gate 46. The $\overline{\text{PRESET}}$ signal is applied to the R labeled input of flip-flop 50, and the $\overline{\text{CLOCK}}$ signal is applied to the C labeled input. The CLOCK signal is applied to the other input of AND gate 52. The output of AND gate 52 is the SHIFT CLOCK signal which also is applied to the other input of AND gate 54 and to the C labeled inputs of generators 26 and 28. The output of AND gate 54 is the COUNT signal. The $\overline{Q}$ output of flip-flop 50 is directed to the C labeled input of a J-K flip-flop 56. A voltage V_{cc} is applied to the J and the K labeled inputs to flip-flop 56. The $\overline{\text{PRESET}}$ signal is applied to the R labeled input of flip-flop 56. The output of flip-flop 56 is taken from the Q labeled output. The output signal is designated OUTPUT VALID.

A pseudo random number generator suitable for use as generator 26 is disclosed in Fig. 2. A string of sixteen D-type flip-flops are connected in serial fashion with the Q output of each flip-flop being connected to the D input of the following flip-flop.

Each of the D-type flip-flops labeled 1-16 is clocked by the CLOCK signal applied to the C labeled inputs. The PRESET signal is applied to each of the R labeled inputs of the sixteen flip-flops to reset the register to an initial condition.

In the preferred embodiment of this invention generator 26 was designed to generate a random number polynomial $x^{16} + x^{12} + x^{11} + x^{9} + x^{8} + x^{4} + x^{3} + x^{2} + 1$.

This was accomplished by Exclusive ORIng each of the first shift register outputs which correspond in number to the exponent of the to be simulated polynomial, for which, x^{16} is simulated by taking the Q output from the D flip-flop numbered 16 and by providing it as an input to an Exclusive OR gate 74. In a like manner, the output of flip-flop 12 is directed to the other input of Exclusive OR gate 74 to simulate the terms $x^{16} + x^{12}$. Correspondingly, the output from flip-flop 11 is combined in Exclusive OR gate 72 with the output from Exclusive OR gate 74 and the output of flip-flop 9 is combined in Exclusive OR gate 70 with the output from Exclusive OR gate 72. Exclusive OR gates 68, 66, 64 and 62, in a like manner, simulate the $x^8 + x^4 + x^3 + x^2$ elements of the polynomial. The output of the Exclusive OR gate 62 is then directed back to the input of an Exclusive OR gate 60 to provide a closed loop path. The last remaining term is derived from the $\overline{Q}$ output of flip-flop 1, which output is labeled G1 and is the output of the pseudo generator 26. The Exclusive OR gate 60 receives as its other input the D signal which data signal is combined on a bit by bit basis with the bits received from the output of Exclusive OR gate 62.

Referring now to Fig. 3, the pseudo random number generator 28 is shown comprised of sixteen serially connected D-type flip-flops labeled 1 through 16. The Q output of each flip-flop is connected to the D input of the succeeding flip-flop. In the preferred embodiment of the invention, pseudo random number generator 26 was designed to implement the polynomial $x^{16} + x^{12} + x^3 + x + 1$. The reset input to each flip-flop is labeled R. The D signal which is received from the Exclusive OR gate 24 is directed to one input of Exclusive OR gate 60. The output of each flip-flop is labeled according to the flip-flop's position in the serial string and corresponds to a bit position in a sixteen bit signal.

Referring to Fig. 4, the output register 32 is illustrated comprised of 16 D-type flip-flops serially connected with the Q output of each flip-flop connected to the D input of the succeeding flip-flop. The CLOCK and PRESET signals are applied to the C and D labeled inputs on each of the 16 flip-flops. The output from output register 32 is taken from the Q output of each of the flip-flops and corresponds to 16 bits of a block of hashed key data.

The hardware implementation of the preferred embodiment of the invention has been set forth above. In operation, the purpose of the apparatus is to randomize (hash) blocks of key data consisting of N bits to a table address space consisting of 2^K entries where K is much less than N. This process is accomplished by utilizing the presettable pseudo-random number generator and the counting circuit in the following steps:

1. Presetting the pseudo-random number generator and the counting circuit to an initialized state.

2. Loading a predetermined count into the counting circuit.

3. Entering key data into the pseudo-random number generator to randomize the key data.

4. Complete the key data randomizing and provide an output valid signal in accordance with the count in the counting circuit.

5. Repeating steps 2, 3 and 4 until all desired key data has been hashed.

Referring to Fig. 1 in conjunction with Figs. 5A-5I, the apparatus is initialized by the PRESET signal (Fig. 5C) being applied to the various preset terminals to set the OUTPUT VALID signal low and to inhibit internal clocks. The constant value K is loaded into the count register 36 by raising the LOAD COUNT signal (Fig. 5D) to a high level and presenting the DATA IN signal

(Fig. 5B) representing the constant value K, bit by bit, on the DATA IN input terminal. This data is then clocked (Fig. 5A) serially into the count register 36 during the LOAD COUNT CYCLE. The value of the constant should be equal to the number of bits in the output register 32 that are used for the hash address. In the preferred embodiment of the invention, K was limited to Hex FF and the number of outputs of the output register 32 were therefore limited to sixteen. After the count register 36 is loaded, activating the PRESET signal (Fig. 5C) again will transfer this value into the down counter 38 and will initialize the pseudo random number generators 26 and 28. A block of key data may then be loaded bit by bit onto the DATA INPUT line to AND gate 20 and count register 36. Activating the LOAD DATA signal (Fig. 5E) enables the key data to be directed to the cross-coupled pseudo random number generators 26 and 28. The LOAD DATA signal going true enables the SHIFT CLOCK signal (Fig. 5F) for the generators 26 and 28 and the serial-to-parallel output register 32. The LOAD DATA going false enables the COUNT signal (Fig. 5G) to the down counter 38. The SHIFT CLOCK signal and the COUNT signal will continue until the terminal count TC (Fig. 5H) is reached in the down counter 38. The terminal count signal will disable the SHIFT CLOCK and the COUNT signal and set the output of flip-flop 56 to indicate an OUTPUT VALID signal. On receipt of the OUTPUT VALID signal, the outputs present on the terminals 1-16 of the output register 32 will be valid hashed data.

In the preferred embodiment of the invention, the outputs from each of the pseudo random number generators is cross-coupled to the input of the other pseudo random generator so as to further scramble (or encode) a DATA IN signal. Exclusive ORing of the output from each of the pseudo random number generators insures a high degree of randomness to the DATA IN signal. The randomized signal is then applied to the output register which register accumulates a selected number of data transitions or data bits, in this case, sixteen bits of data and outputs the data in blocks of sixteen.

From the aforementioned description of the preferred embodiment of the invention, it can be seen that there is an advantage in that a uniform distribution of key mappings into the table address space as the number of keys becomes large relative to the size of the table address space is provided and that the apparatus removes any properties which the original keys may have, such as determinate relationship to each other. In addition, the apparatus is independent of the key length.

CLAIMS:

1. Apparatus for hashing key data, characterized by pseudo-random bit generating means (22, 24, 26, 28, 30) responsive to key data signals applied thereto to generate a randomized output signal, output register means (32) adapted to receive said randomized output signal and control means (36-56) adapted to provide an output valid signal indicative of valid hashed key data being available at an output of said output register means (32).

2. Apparatus according to claim 1, characterized in that said pseudo-random bit generating means (22, 24, 26, 28, 30) is responsive to shift signals and in that said control means is responsive to the provision of a predetermined number of said shift signals to control the production of said output valid signal.

3. Apparatus according to claim 2, characterized in that said pseudo-random bit generating means includes first and second pseudo-random generators (26, 28) and combining means (22, 24) adapted to combine said key data signals with each of first and second output signals of said first and second pseudo-random generators and to apply the resulting first and second combined signals to the second and first pseudo-random generators (28, 26) respectively.

4. Apparatus according to claim 2, characterized in that said control means includes: a counter device (38) settable to a load count dependent on the number of elements included in said hashed key data, said counter device (38) being responsive to the application of selected ones of said shift signals to provide a terminal count signal; and logic means (44-56) responsive to said terminal count signal to provide said output valid signal.

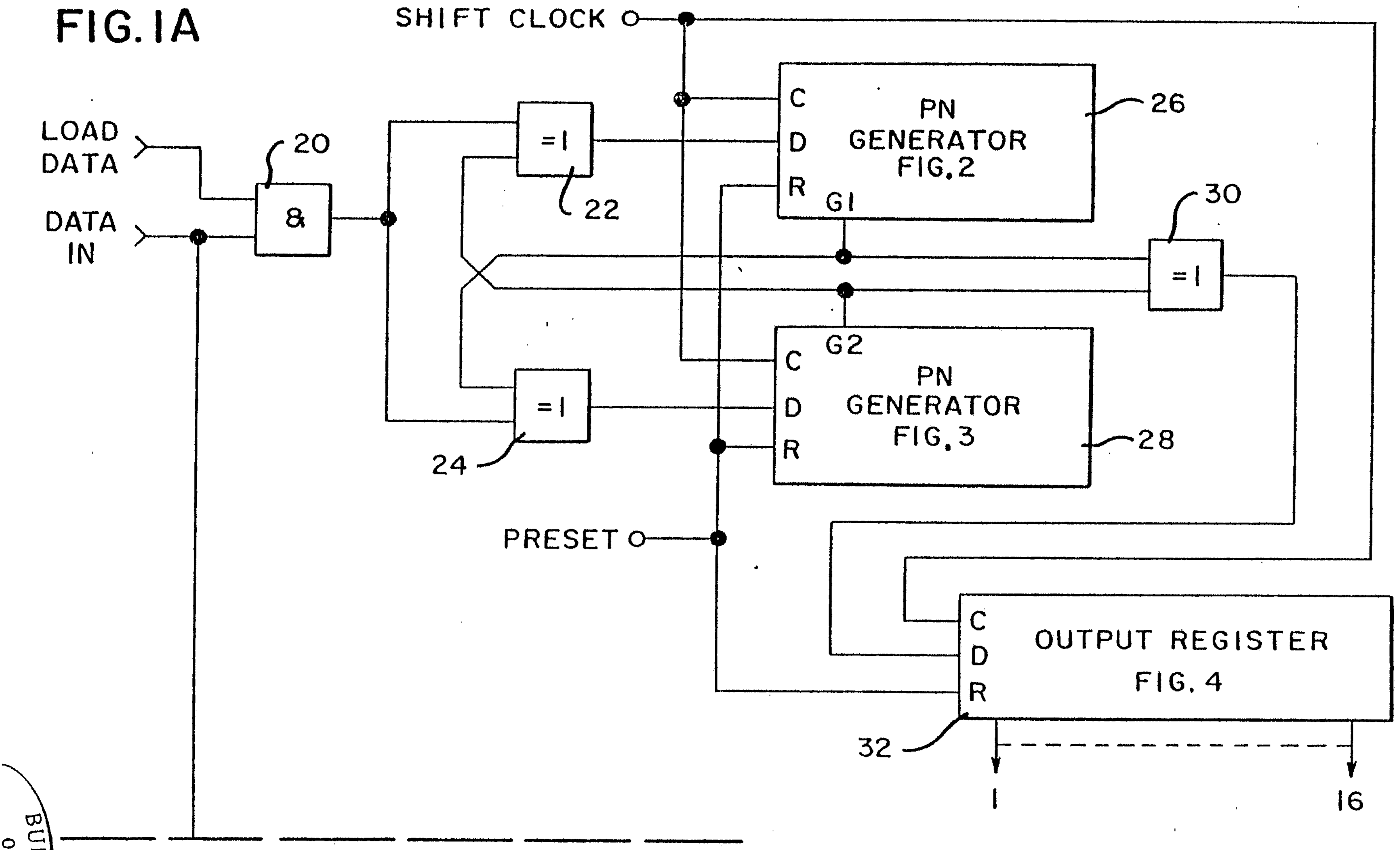
5. Apparatus according to claim 4, characterized in that said control means includes a count register (36) arranged to be supplied with signals representing said load count, said count register being coupled to said counter device (38) and adapted to transfer said load count thereto in response to a control signal (PRESET).

6. Apparatus according to claim 3, characterized in that said first and second pseudo-random generators (26, 28) include respective first and second shift registers operable in response to said shift signals.

7. A method for hashing key data, characterized by the steps of: (a) presetting pseudo-random bit generating means (22, 24, 26, 28, 30) and a counting device (38) to an initialized state; (b) setting a predetermined load count into said counting device (38); (c) applying key data to said pseudo-random bit generating means (22, 24, 26, 28, 30); and (d) generating a randomized output signal by applying shift signals to said pseudo-random bit generating means in accordance with the loaded count.

8. A method according to claim 7, characterized by the steps of repeating steps (b), (c) and (d) until all desired key data has been hashed.

FIG.1A


SHIFT CLOCK
LOAD DATA
DATA IN
20
&
=1
22
=1
24
C
D
R
PN GENERATOR FIG.2
G1
26
30
=1
G2
C
D
R
PN GENERATOR FIG.3
28
PRESET
C
D
R
OUTPUT REGISTER FIG. 4
32
1
16

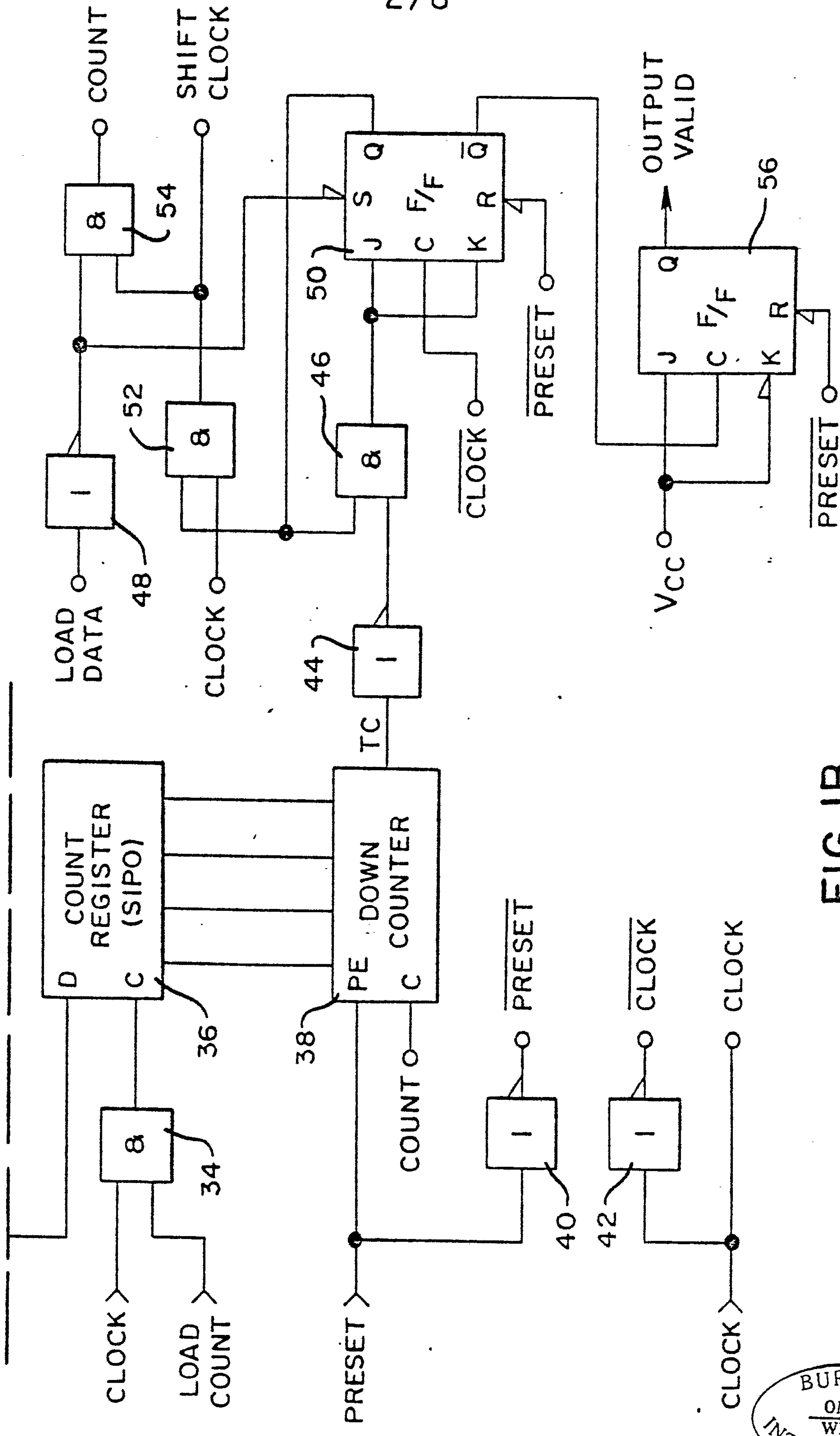
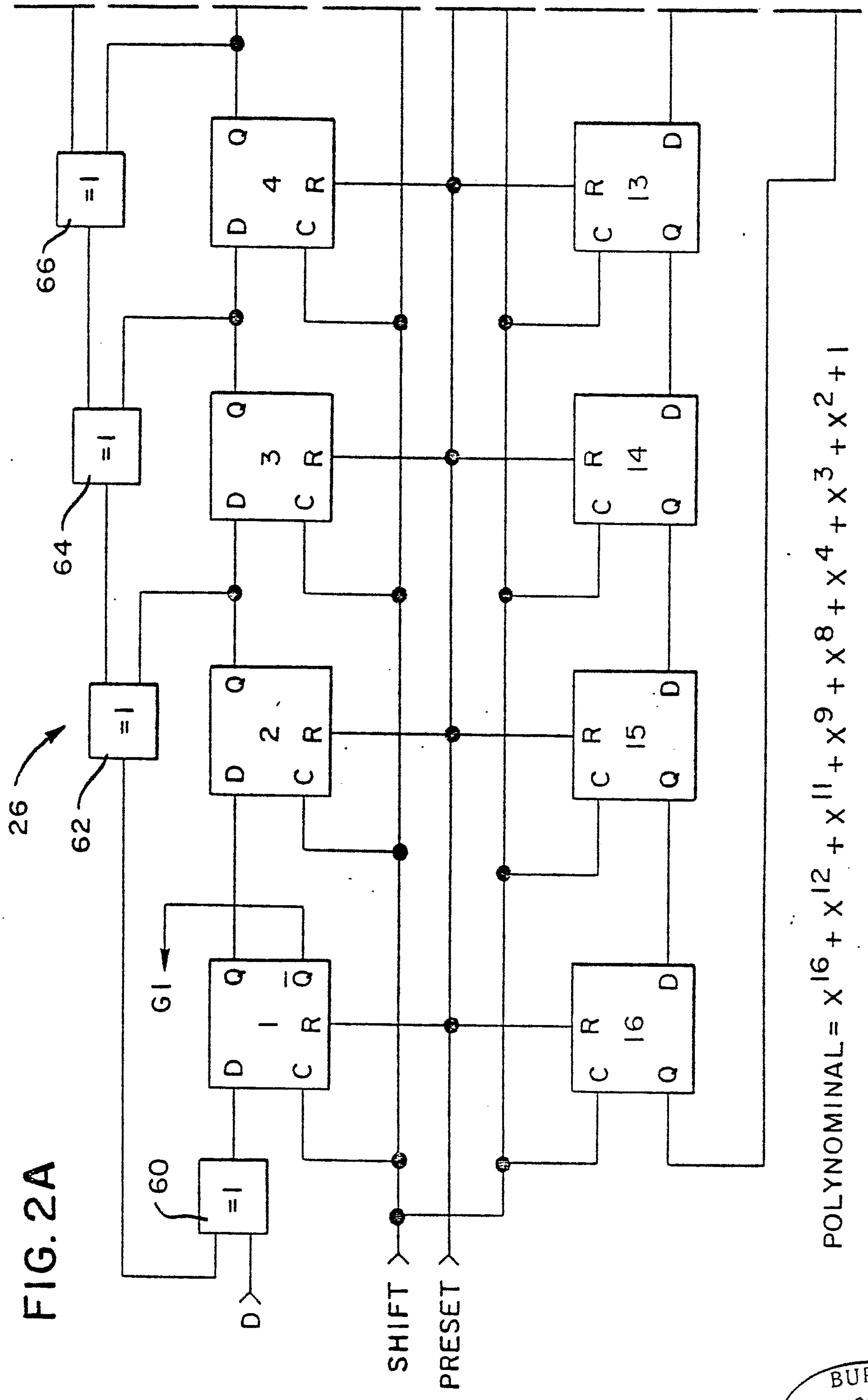

COUNT
SHIFT CLOCK
OUTPUT VALID
54
50
46
52
48
44
56
&
1
F/F
J
S
Q
C
K
R
$\overline{Q}$
PRESET
$\overline{CLOCK}$
$\overline{PRESET}$
V_{CC}
LOAD DATA
CLOCK
TC
COUNT REGISTER (SIPO)
D
C
36
DOWN COUNTER
PE
C
38
COUNT
34
40
42
$\overline{PRESET}$
$\overline{CLOCK}$
CLOCK
CLOCK
LOAD COUNT
PRESET
CLOCK

FIG. 1B

FIG. 2A


26
60
62
64
66
61
D
Q
C
R
Q̄
1
2
3
4
13
14
15
16
=1
D
SHIFT
PRESET
POLYNOMINAL = X^16 + X^12 + X^11 + X^9 + X^8 + X^4 + X^3 + X^2 + 1

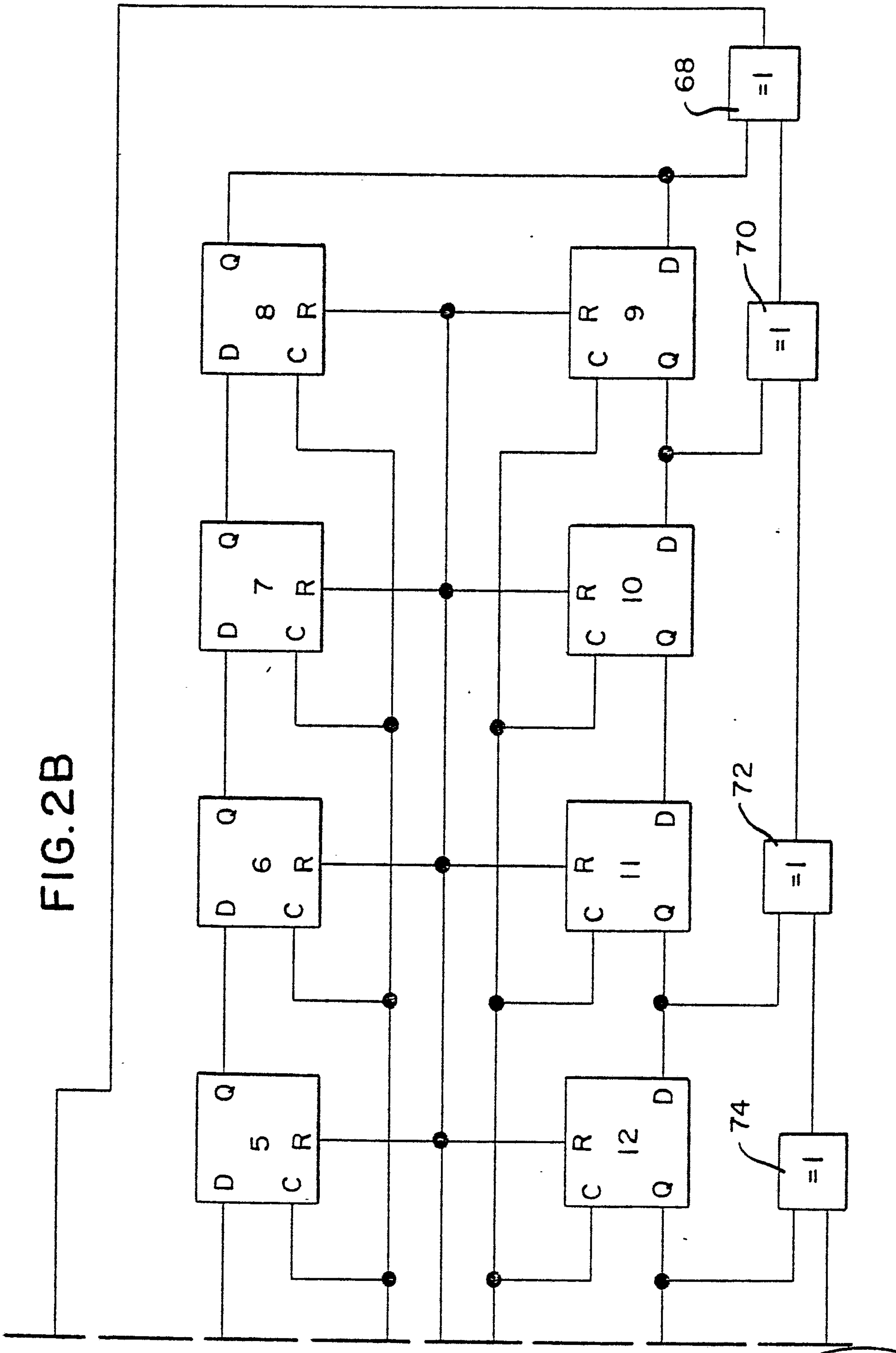
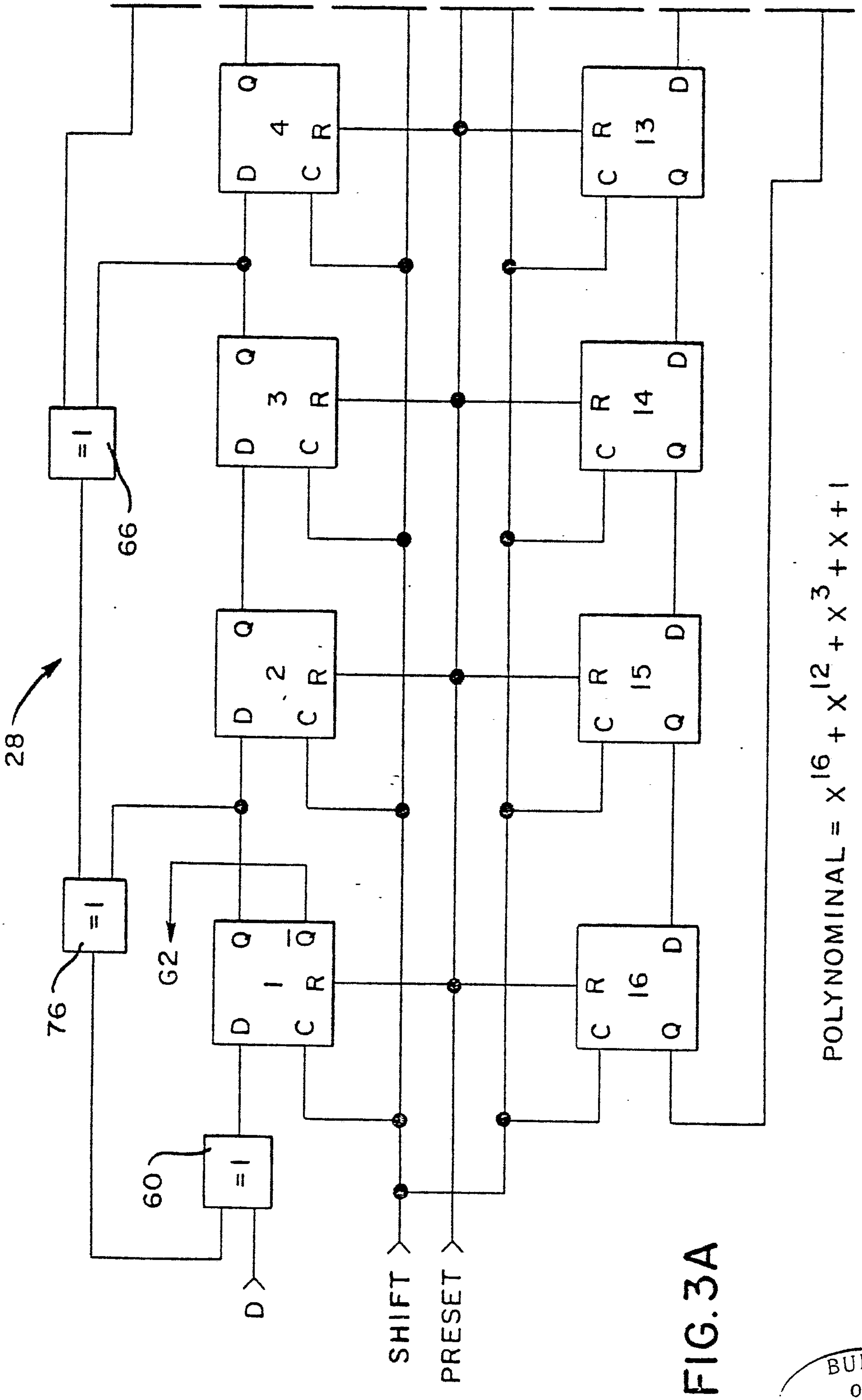

68
=1
70
=1
72
=1
74
=1
D Q
8
C R
D Q
7
C R
D Q
6
C R
D Q
5
C R
R D
9
C Q
R D
10
C Q
R D
11
C Q
R D
12
C Q

FIG.2B



POLYNOMINAL = $X^{16} + X^{12} + X^3 + X + 1$

FIG. 3A

FIG. 3B

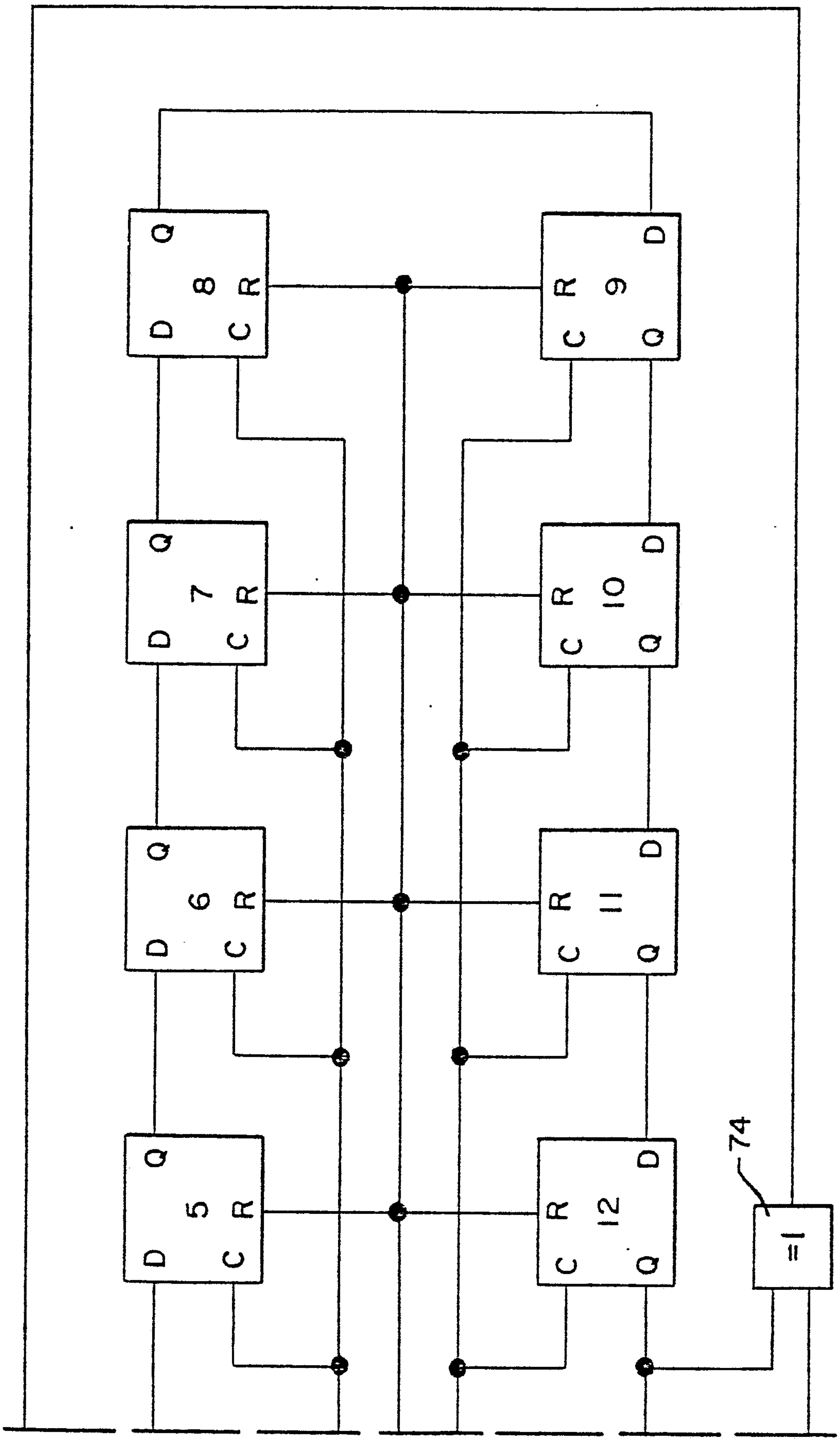


FIG. 4

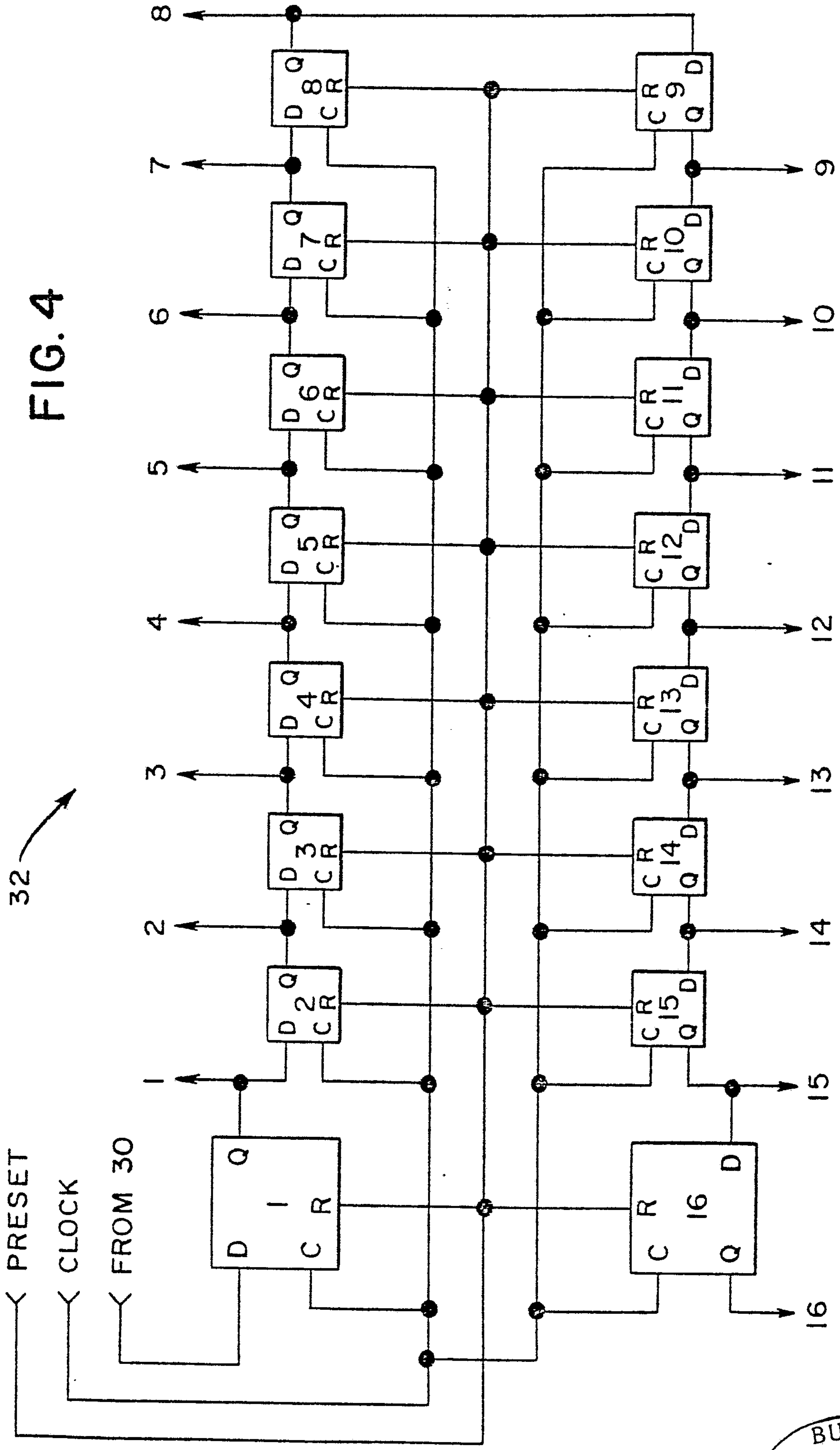
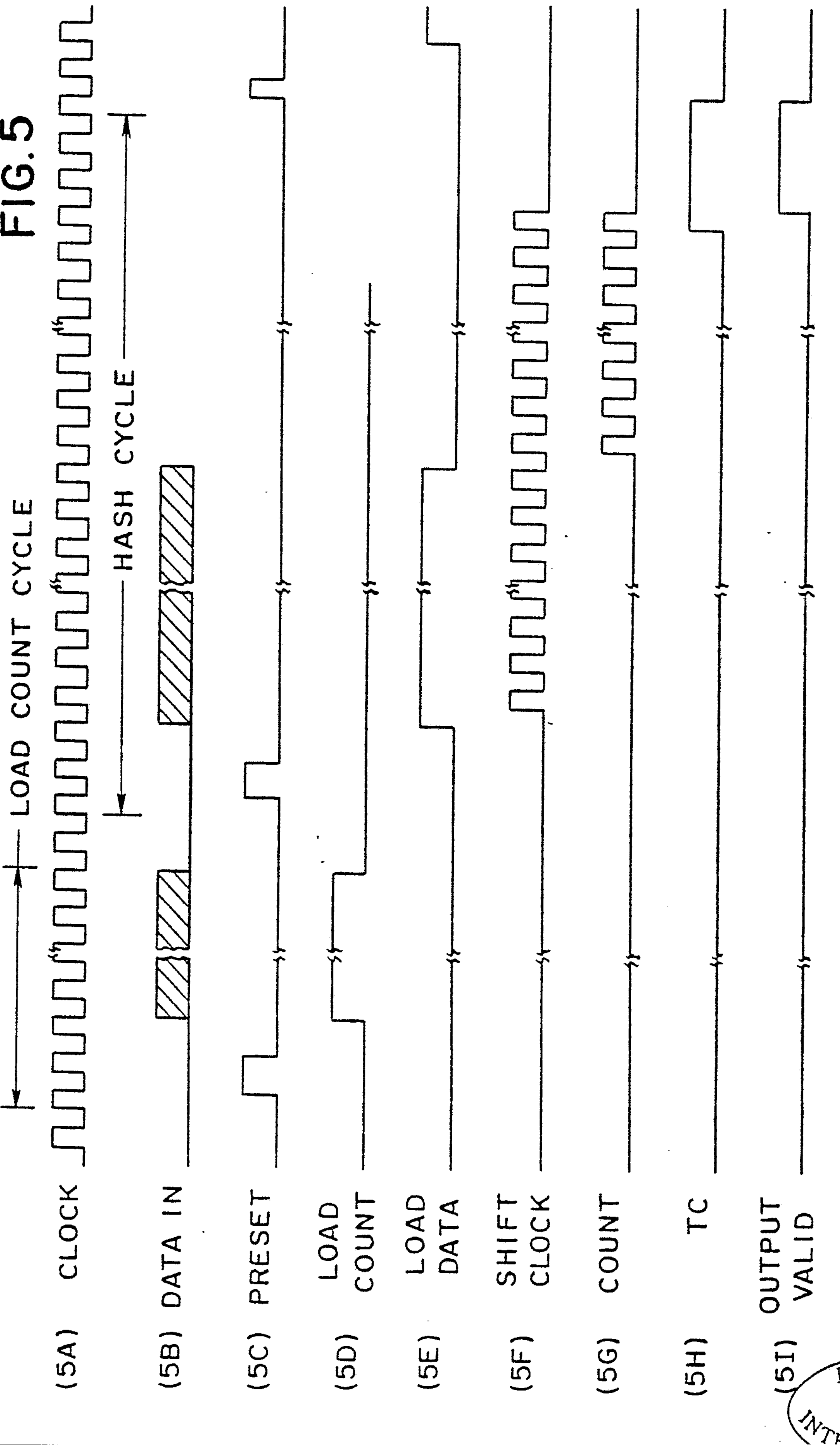


FIG.5


LOAD COUNT CYCLE
HASH CYCLE
(5A) CLOCK
(5B) DATA IN
(5C) PRESET
(5D) LOAD COUNT
(5E) LOAD DATA
(5F) SHIFT CLOCK
(5G) COUNT
(5H) TC
(5I) OUTPUT VALID

INTERNATIONAL SEARCH REPORT

International Application No PCT/US80/01597

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [3]

According to International Patent Classification (IPC) or to both National Classification and IPC

INT. CL.^{3} G 06 F 3/00

U.S.CL. 364/200

II. FIELDS SEARCHED

Minimum Documentation Searched [4]

Classification System	Classification Symbols
U.S.	178/22; 179/1.5R 235/92DE, 92 SH 364/200, 900

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [5]

III. DOCUMENTS CONSIDERED TO BE RELEVANT [14]

Category *	Citation of Document, [16] with indication, where appropriate, of the relevant passages [17]	Relevant to Claim No. [18]
X	US, A, 4,157,454 BECKER Published 05 June 1979	1-8
X	US, A, 4,115,657 MORGAN Published 19 September 1978	1-8
X	US, A, 4,112,487 NUTTER Published 05 September 1978	1-8
X	US, A, 3,784,743 SCHROEDER Published 08 January 1974	1-8
X	US, A, 3,691,472 BOHMAN Published 12 September 1972	1-8

* Special categories of cited documents: [15]

"A" document defining the general state of the art

"E" earlier document but published on or after the international filing date

"L" document cited for special reason other than those referred to in the other categories

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but on or after the priority date claimed

"T" later document published on or after the international filing date or priority date and not in conflict with the application, but cited to understand the principle or theory underlying the invention

"X" document of particular relevance

IV. CERTIFICATION

Date of the Actual Completion of the International Search [2] 27 March 1981	Date of Mailing of this International Search Report [2] 02 APR 1981
International Searching Authority [1] ISA/US	Signature of Authorized Officer [20] J. M. Thesz Joseph M. Thesz, Jr.