PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : H04L 12/56, 12/18	A1	(11) International Publication Number: WO 92/08306 (43) International Publication Date: 14 May 1992 (14.05.92)

(21) International Application Number: PCT/GB91/01909

(22) International Filing Date: 31 October 1991 (31.10.91)

(30) Priority data:
90312134.1 6 November 1990 (06.11.90) EP
(34) Countries for which the regional or international application was filed: GB et al.

(71) Applicant *(for all designated States except US):* HEWLETT-PACKARD COMPANY [US/US]; 3000 Hanover Street, Palo Alto, CA 94304 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)* : WAINWRIGHT, Nicholas, John [GB/GB]; 36 Northumberland Road, Bristol BS6 7BD (GB).

(74) Agent: COKER, David, Graeme; Hewlett-Packard Limited, IP Section, Building 2, Filton Road, Stoke Gifford, Bristol BS12 6QZ (GB).

(81) Designated States: JP, US.

Published
With international search report.

(54) Title: MULTICAST SWITCH CIRCUITS

(57) Abstract

A multicast switch fabric is described wherein blocking of desired signal paths is handled by storing a cell to be transmitted through the fabric at a buffer at an input port to the fabric and attempting to retransmit the cell through the fabric until the cell is successfully transmitted, or a maximum number of retries at transmission is reached. Each switch element in the fabric has a block register which stores information as to whether or not a request path including that element is blocked, the block signals being fed back to the input port to indicate whether or not cell transmission has been successful. A 2 x 2 switch element is described wherein header information from the cell is decoded in two subfields, one field determining the number of copies to be replicated in the switch and a second field determining if the cell is to be fully replicated at each subsequent stage of the fabric.

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	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland			SN	Senegal
CI	Côte d'Ivoire	KR	Republic of Korea	SU+	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
CS	Czechoslovakia	LK	Sri Lanka	TG	Togo
DE*	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		

+ Any designation of "SU" has effect in the Russian Federation. It is not yet known whether any such designation has effect in other States of the former Soviet Union.

MULTICAST SWITCH CIRCUITS

Technical Field

This invention relates to multicast switch circuits for use in packet switches operating to transfer data in an asynchronous transfer mode (ATM) network.

With the arrival of optical fibre communication links, broadband packet networks are required to handle and transfer packets of information such as data, voice, and video at high speed from multiple inputs to multiple outputs. Multicast packet switches can be employed in such networks and may consist of two parts - a multicast fabric wherein the number of copies of a packet are created from a single input packet, and a routing fabric wherein the required number of packets as produced by the multicast fabric are received and directed to the correct output addresses.

Background Art

Various multicast packet switches have been proposed in the past, for example those by J.S.Turner (IEEE Transactions on Communications, 36, no. 6, June 1988, pp.734-743) and T.T.Lee (IEEE Journal on Selected Areas in Communications, 6, no. 9, December 1988, pp.1455-1467). Such switches are normally modular in design and are intended for implementation on a large scale, coupled for example in a delta network, so as to enable large numbers of inputs/outputs to be handled.

A problem with such multicast fabrics is how to handle contention at nodes within the fabric, where collisions can occur between cells (packets) contending for the same paths within the fabric. Blocking arises when a cell is unsuccessful in gaining control of a suitable path through the fabric. A means must be provided for enabling the cell to retry transmission so as to complete its passage through the switches. Lee (ibid) describes a particular form of non-blocking multicast circuit, whereas Turner (ibid) handles blocking by providing buffers for the cells at the nodes within the multicast fabric and which hold the cells until a successful retry in transmission through further nodes in the fabric can be made.

The present invention is concerned with switches for multicast fabrics which handle cell replication and blocking in a novel way. They can be produced cheaply in small gate arrays, and therefore are suitable for production in modular form to enable multi-input/output packet switches to be produced.

Disclosure of Invention

According to the invention there is provided a multicast switch element suitable for connection to a plurality of similar such elements in a matrix to form a multicast switch fabric, said element having a plurality of inputs and a plurality of outputs and a switch therebetween for passing through to the outputs cells of information received at the inputs and for replicating copies of the cell information in accordance with instructions associated with the cell information, and means associated with the switch for decoding said instructions in two sections: one section determining the number of times the cell information is to be replicated within the switch, and the other section determining whether or not the cell is to be replicated fully in any subsequent switch elements to which the element may be connected in the multicast switch fabric.

According to the invention there is also provided a multicast switch fabric having an M x M matrix of connected switch elements as specified in the preceding paragraph, an input port and an output port associated with each input and output respectively of the fabric, each said input port having an input buffer for storing a cell for transmission through the fabric and means for retaining any such cell in said buffer until cell transmission is successful.

Preferably the switch element is constructed as a small modular unit, e.g. as a 2 x 2 switch, which can be coupled with many similar elements to form a matrix such as a delta network.

In accordance with the invention the incoming cell is buffered at an input port of the multicast fabric where it remains until the cell can be successfully transmitted through the network of switch elements, or until a maximum number of retries at transmission is reached. A succession of attempts is made to set up the appropriate path through the fabric. If this cannot be achieved in any one cycle (as a consequence of blocking occurring) then a blocking signal is fed back to the input port from the blocked node in the network. Each switch element preferably has its own blocking register to store information as to whether it, itself, is blocking an incoming cell or if there is another switch element downstream in the attempted transmission path which is blocking the transmission.

Brief Description of Drawings

Preferred features of the invention will now be described by way of example, with reference to the accompanying drawings, in which:-

Figure 1 is an illustration of a simple multicast switch fabric according to the invention in generalised form;

Figure 2 illustrates the format of a packet or cell as employed in the present invention;

Figure 3 illustrates generally how a cell, as in Figure 2, is switched through the multicast switch fabric of Figure 1;

Figures 4 and 5 illustrate cell blocking and cell resubmission in the multicast switch fabric according to the invention;

Figure 6 is a schematic diagram in block form of a multicast switch element according to a first embodiment of the invention; and

Figures 7 to 21 describe in more detail (down to gate level) a second embodiment of the invention.

Best Mode for Carrying Out the Invention, & Industrial Applicability

Referring to Figure 1, the multicast fabric is an input-buffered delta network comprising S stages (3 stages as illustrated) of M x M multicast switch elements 2, where M = 2^m, m being an integer - 1 as illustrated. At each input port, an input port controller 4 buffers incoming cells and submits them to the switch fabric in sequence. The latter operates cell synchronously so that all controllers are synchronised to present cells to the fabric simultaneously. The switch elements 2 are unbuffered and operate in two phases: a set-up phase during which the switch elements interpret cell header information and try to set up a path from the input ports to one or more output ports, and a data phase during which the data is transferred across the previously established paths to the switch fabric outputs. The data paths between the switch elements are shown in solid lines, whereas the feedback paths (the blocking signals) are shown by dotted lines.

Referring to Figure 2, each cell or packet has a header 6 which includes a copy control field which defines the number of times the cell must be duplicated as it passes through the switch fabric. Each element 2 in the switch fabric interprets the copy control fields of the cells that arrive at its inputs, sets up a path from the input to one or more of its outputs, and modifies the copy control field of each cell that appears at its outputs to reflect the number of times that it must be duplicated by the following stages in the switch fabric. The copy control field consists of the desired number of copies, C_s, encoded as s pairs of subfields, $[R_s,P_s]$. As a cell proceeds through the switch the copy control field is stripped from the cell, stage by stage, to ensure that the subfield for a given stage is always at the head of the cell when it arrives.

Consider a switch element at stage s. Each of its outputs are connected to a switch element at stage s-1. The number of switch fabric outputs that can be reached from each of its outputs is M^s. In the multicast switch this represents the maximum number of copies that can be made of a cell appearing at the output of a switch element by the following stages of the fabric. On entry to a switch element at stage s, the R_s (replicate) subfield defines the number of cells that must be fully replicated at its outputs (i.e. cells that should be copied to every switch fabric output that may be reached from the stage to its right). These cells will produce a total of R_sM^s cells at the outputs of the switch fabric.

The P_s (pass-through) subfield defines whether a copy of the cell should also be passed through with its cell header unchanged. This is in addition to the cells that appear at the switch element outputs for full replication by subsequent stages.

On entry to a switch element at stage s:

$$C = R_{s-1}M^{s-1} + R_{s-2}M^{s-2} + \ . \ . \ . + R_1M + R_0$$

$P_s = 1$ if the replicate field for any of the subsequent stages, R_{s-1} to R_0, is non zero, and is set to zero otherwise.

Figure 3 illustrates how a cell is routed through a three stage network of 4 port elements (for simplicity, only the top half of the network is shown). Eleven copies of Cell 1 are to appear at the switch fabric outputs. At stage 2 (the first stage) $R_2 = 0$, $P_2 = 1$, and the cell is passed through unchanged, losing its copy control field for stage 2 on the way. At stage 1, $R_1 = 2$ and $P_1 = 1$, the cell is again passed unchanged and an addition two copies of the cell are produced, each with $R_0 =4$ to ensure full replication in the final stage. At the last stage, a total of eleven copies of the cell are produced. In the same example, cell 2 requires 17 copies. At the first stage, $R_2 = 1$ and $P_2 = 1$, producing one fully replicated cell that goes on to produce a total of 16 copies plus one cell that passes through the rest of the fabric to produce the 17th cell. Note that after the first stage, the cells that are to be fully replicated have the subfields in the copy control field changed (or "forced") to 4,0, to ensure 4 copies are produced at each subsequent switch element.

The multicast switch fabric as described is a blocking network where collisions can occur between cells contending for the same paths within the fabric. Figures 4 and 5 explain how the circuit of the invention handles such blocking and how resubmission of blocked cells arises to enable a subsequent successful transmission to occur. Referring to Figure 4, cell 1 successfully produces 11 copies, but this time cell 2 starts at a different input to the fabric and its path results in collisions, C, with cell 1 at 11 of the outputs that it needs in the final stage. If the fabric only allowed cells that were successful on all paths to produce output cells, the performance of the fabric would be poor. This problem is addressed in a buffered fabric such as Turner (ibid) by storing cells within the fabric where the blocking occurred and retrying the cell when the blocked output becomes clear. Without buffering within the network, some other technique must be employed to ensure reasonable performance.

In accordance with the invention the performance of the unbuffered multicast fabric is improved by allowing any successful path to produce output cells even when some branches of the paths are blocked. The state of unsuccessful paths is preserved within the fabric. Any cell that is blocked on one or more paths is resubmitted by the input port controller to allow previously blocked paths to complete. This maximises the number of cells that can be output in any cycle. This is illustrated in Figure 5.

The input port controller can limit the maximum of times a cell is submitted to the fabric. An extra flag, the L, last-attempt flag can be added to the copy control field to indicate that a cell is being submitted for the last time. When a cell with the L flag set passes through the fabric, the switch elements release all paths for that cell, blocked or otherwise.

The multicast switch element feeds the results of contention for outputs back to the preceding stage. The feedback path indicates whether a cell presented at a switch element input has been blocked on any of its required paths. Blocking can occur either internally within the switch element or at a subsequent stage in the fabric.

A first embodiment of the invention is illustrated in Figure 6 as a 2 x 2 switch element. The two data inputs, INPUT 0 and INPUT 1, can provide outputs at OUTPUT 0 and/or 1, thus either passing a cell straight through without replication or by creating one extra copy of the cell. The multicast fabric consists of cell header decoders 8, requestors 10, cell header overwriters 12, selectors 14, and demultiplexers 16.

The circuit operates as follows. During the set-up phase, the demultiplexers 16 at each output are configured to select a cell from one of the inputs. During the data phase, the cell is transferred to the output ports and feedback information ripples back to the input port controller (not shown).

The requestor 10 receives R_s and P_s subfields at the head of the copy control field from decoder 8 and determines which output the cell should be sent to. This information is saved in a request register within the requestor. The requestor attempts to reserve each output that it needs by signalling a request to each of the corresponding selectors 14.

The selector 14 arbitrates between conflicting requests from different input ports for the same output. Once selected, an output port stays assigned to the same input port until all copies of the cell have been made by subsequent stages. This is indicated by the feedback signal from the following stage indicating unblocked. The selector uses round-robin arbitration to ensure fairness. After completion of a cell transmission, the input that was last selected is given lowest priority during the next set-up phase.

The remainder of the cell, including the copy control field for subsequent stages, is passed through the switch element once the data paths have been set up. The copy control fields of each of the copies of the cell that are to be fully replicated by subsequent stages are adjusted by the cell header overwriter 12 so that R = M, P = 0.

The feedback signal becomes valid after the set-up phase for the entire fabric is complete. This signal ripples back from the final stage to the switch fabric input port controller. In the switch elements of the final stage (which never receive a blocked feedback input as they have no subsequent stage), each selector at the switch element outputs issues an ack signal to each of the requestors that were contending for it to indicate which was allocated to the corresponding output port.

The requestors at the switch element inputs receive an ack signal from each of the selectors in the switch element outputs and save these in a blocked register. The feedback signal from the switch element indicates blocked if any of the switch elements necessary for the cell have not yet been allocated (i.e. if any bit that is set in the request register is indicated as blocked by the blocked register).

At the preceding stage, the incoming feedback signal indicates blocked if any path on stages to its right has not yet been completed. In this case, the requestor that had successfully reserved the output will still receive a blocked indication on the corresponding ack signal from the selector. It will therefore pass on a blocked feedback signal to the stage that precedes it. If the incoming feedback signal indicates not-blocked, the selector can release the output port at the end of the current cell and participate in the next set-up phase.

A second embodiment of the invention is illustrated in more detail in Figures 7 to 22. This again is a 2 x 2 multicast switch element, as in Figure 6, and although it shares some common features with the latter embodiment, there are also differences. For example, the copy control field of each cell contains no "L" or last-attempt flag. In this embodiment the number of unsuccessful attempts at transmission is counted at the input port. When the count decrements to zero, the switch is reset to abort cell tranmission. A further attempt at retransmission is then subsequently made in the next overall transmission cycle.

Figure 7 shows the multicast switch element 102 generally together with its corresponding input and output ports, 104 and 106. The circuit 102 receives cells on data input lines DIN0 and DIN1 together with a corresponding request for transmission signals on lines RQIN0 and RQIN1. Indications of whether succeeding stages are blocked are received from the output port 106 (BL0 and BL1) and these are passed back to the input ports. A blocking signal will be passed back if either (a) a blocked signal has been received from a succeeding stage or (b) if a block is created within the stage itself. The circuit 102 also receives, and transmits onwardly, a reset signal (input port 104, "RESET") in the event of a last-attempt at transmission, as described above. The reset clears any blocks for the cell whose transmission is aborted.

The input and output ports 104,106 are not part of the multicast circuit per se and the output port 106 is essentially devoid of any circuitry, being basically a straight-through connection to its output. Although the input port is not shown in more detail, it includes an input buffer which receives the incoming cell and stores it until transmission through the multicast circuit is successful, a retry counter which generates the reset signal to abort transmission after a given number of blocked attempts, and a look-up table which appends the copy control field to the cell appropriate for the passage of the latter through the multicast circuit.

The multicast circuit 102 is shown in more detail in Figure 8, and in succeeding Figures 9-21. In these Figures, where standard components are shown conventionally illustrated, these will not be explained in more detail. The switch proper which receives data on one or other of two inputs and transmits across to one or other or both of two outputs is a cross-bar switch 108. A similar cross-bar switch 110 correspondingly receives and transmits the request signals. The circuit 102 also includes a request delay circuit 112, a fairness register 114, a reservation register 116, a block register 118, a forcer 120, and an overall control circuit 122.

In this embodiment, the copy control field (CCF) of each cell is 2 bits long, one bit (NC) determining the number of copies of the cell to be replicated within the 2 x 2 switch and the other (FULL) dictating whether or not there is to be full replication of the cell at each subsequent stage in the multicast fabric. Thus:-

NC	FULL		
0	0	:	no copy to be made, no full replication subsequently
0	1	:	no copy, full replication subsequently,
1	0	:	1 copy, no replication on one port, full replication on the other
1	1	:	1 copy, full replication on both outputs subsequently.

A feature of the present switch element is that if full replication is subsequently desired at any output, then "11" is forced into the copy control field of the transmitted outgoing cell.

The request delay circuit 112 is shown in Figure 9 and consists of eight D flipflops FD. The request delay circuit delays the request signal by two clock cycles so as to "strip off" the two bit CCF signal for this particular stage of the switch fabric from the cell signal. The two bit signal is latched to provide NC and FULL signals for both inputs 0 and 1 (NC0, FULL0; NC1, FULL1). Thus, with an S stage multicast fabric and assuming the 2 x 2 switch is the first stage (stage S) in the fabric then the CCF of the incoming cell would be

(NC_s, $FULL_s$; NC_{s-1}, $FULL_{s-1}$ .... NC_0, $FULL_0$).

In this example the request delay circuit would strip off NC_s, $FULL_s$ and store the values thereof in the appropriate latches.

Cross-bar switches 108,110 are shown in more detail in Figure 10. The inputs it is desired to switch in the case of switch 108 are the cells or packets themselves (information plus CCF) whereas in the case of switch 110 it is the request signals that are switched. To whichever output port an incoming cell is switched to in switch 108, its corresponding request signal is switched similarly to the same port synchronously through switch 110. Because of the delay introduced by the request delay circuit 112, the cell is transmitted through switch 108 devoid of that part of the CCF for the stage in question. In other words, to return to the example above, (NC_s, $FULL_s$) is lost and the CCF is clocked through as (NC_{s-1}, $FULL_{s-1}$; NC_{s-2}, $FULL_{s-2}$ ....) so that the correct two subfields of the CCF are transmitted at the front of the head of the cell for the next stage in the fabric.

Both cross-bar switches are switched in accordance with the SEL0 and SEL1 signals as follows:-

SEL0	SEL1	I/P O/P	
0	0	0---->0 1---->1	straight through
0	1	0---->0 0---->1	2 copies of I/P 0
1	0	1---->0 1---->1	2 copies of I/P 1
1	1	0---->1 1---->0	crossover

The fairness register 114 is shown in more detail in Figure

11. This consists of a pair of D flipflops which store 1 bit per output, determined by which input used the output during the last data cycle. For example, the flipflop providing signal TOUT0 stores 0 if I/P 0 used O/P 0 last and stores 1 if I/P 1 used O/P 0 last. The result of such store is for use in any arbitration necessary to decide which input gains a particular output in the event of collision of the two incoming signals in the next set-up phase.

The reservation register 116 is shown in Figure 12, and consists of 4 D flipflops. Subsequent to any arbitration made between colliding requests (see below) this stores the granted requests for the cross-bar switches. The topmost flipflop stores a signal relevant to whether or not O/P 0 can be reserved for I/P 0 (i.e. notation "0.0"). The other three flipflops similarly store requests for I/P 0 - O/P 1 (0.1), etc. Each flipflop is set if, as a result of arbitration, that particular path can be granted (i.e. it blocks the path so that it cannot be granted again during the set-up phase). Each flipflop is cleared if the output is not blocked.

The block registers 118 are shown in Figure 13 and are similar to the reservation registers 116. Essentially the block registers store the reverse information to that of the reservation registers: each stores for a given input, which outputs that input is blocked on. Thus, the topmost flipflop stores the information concerning whether or not I/P 0 is blocked on O/P 0 (0.0). The other flipflops have similar notation. The flip-flops are cleared on the reset line from the relevant reset signal. The latter originates from an input port in the event of a "last try attempt". Blocking is cleared, whilst the cell transmission is aborted.

Forcer 120 is shown in Figure 14. If, as a result of decoding the CCF, the cell is to be fully replicated in succeeding stages, the forcer overwrites NC=1, FULL=1 into the leading subfield of the CCF of the outgoing cell. The forcing signals FORCE 0 and FORCE 1 are obtained from a force signal generator within the control circuit 122 described below.

The control circuit 122 is illustrated in Figure 15. It includes a request decoder 124, a request arbitrator 126, a reservation register controller 128, a block decoder 130, a reset signal decoder 132, a force signal generator 134, and a finite state machine circuit 136.

The request decoder 124 is shown in Figure 16. This converts the request from the two input ports (RQ 0 and RQ 1) into four request signals (e.g. RQ 0.0 indicates I/P 0 is making a new request for O/P 0) depending upon which outputs the cell on a given input is to be transmitted to. If an input is previously blocked on a given output, signifying that this is a retry for this particular transmission path, a new request is not generated as it will already exist.

The request arbitrator 126 is illustrated in Figure 17 and makes a decision which input should get a particular output in the event both inputs request the same output. It arbitrates based on the values stored in the fairness register 114 and, in the event of a collision, it grants signals to create the opposite request to that granted in the last transmission cycle (as stored by the fairness register 114). Thus, if I/P 0 requests O/P 0 and the reservation register 116 for O/P 0 is not set and if I/P 1 has not requested O/P 0, then I/P 0 is granted O/P 0. Output S0 is then set such that I/P 0 will be granted O/P 0. If both I/P 0 and I/P 1 request O/P 0, then which request is granted is determined by the inputs T0 and T1 from the fairness register 114. This will cause S0 to adopt a state enabling the reverse input-output path to be granted to that of the last transmission cycle. The reservation register controller 128 is shown in Figure 18. This creates the signals to set up the reservation register 116 based upon the results of arbitration in arbitrator 126, the requests received from the request decoder 124 and blocking signals from the succeeding stages.

The block decoder 130 is shown in Figure 19. This generates the blocking signals to be stored in the block register 118. It receives the requests from the request decoder 124, the result of arbitration in arbitrator 126, and the blocking signals from subsequent stages. The blocking signals (B 0.0 for I/P 0 blocked on I/P 0 etc.) can be generated in two ways. An input can be blocked on a given output if it made a request for that output (the output not being blocked by subsequent stages) but it failed arbitration and therefore the request was not granted in the stage. Alternately, an input can be blocked on a given output if it made a request for that output (this request was granted within the stage) but it was blocked by the blocking signal from subsequent stages. If an input is blocked it will retry on the next cycle.

The reset signal decoder 132 is illustrated in Figure 20. The RESET process occurs when a cell has been blocked more times than the input port 104 cares to retry. The reset signal is generated from the retry counter in the input port. The switch element uses the signals stored in the block register 118 to determine which outputs are blocked and then uses the reset signal to send the reset down the appropriate path. This clears the path in the block and reservation registers 118, 116 and also transmits the reset signal on to the next stage in the fabric.

The force signal generator 134 is shown in Figure 21 and decodes the CCF to determine whether the output CCF for the outgoing cell should be forced to 11 (i.e. full replication). It is timed by a FORCE timing signal from the finite state machine circuit 136 for the first two bits of the outgoing cell. Both are driven to "1" by forcer 120 thus giving "11" at the head of the CCF.

The finite state machine circuit 136 is not shown in more detail but consists of a pulse generating circuit to generate the various clock signals to synchronise operation of the other components.

The switch operates as follows. Assume that a multicast fabric containing a delta array of the 2 x 2 switches as just described is arranged in S stages, and that a cell arrives at one of the input ports of the fabric. The cell will contain header information dictating the number of cell copies required. From the look up table at the input port the appropriate copy control field consisting of S subfields NC, FULL will be prepended to the cell. The retry counter at the input port is set.

The set-up phase for the fabric now commences. The cell is passed to the first stage 2 x 2 switch (such as in Figures 8 onwards). The cell header information for this stage (stage S), i.e. NC_s, $FULL_s$, is stripped from the header by delay 112 (Figure 9) and this is passed to the request decoder 124 (Figure 16) which decodes from NC_s, $FULL_s$ any new requests from the two inputs for the two outputs. This information is sent to the reservation register controller 128 (Figure 18) and thence stored in the reservation register 116 (Figure 12). The requests for outputs are also sent to the request arbitrator 126 (Figure 17) to determine whether or not the requests can be granted based on fairness (the fairness register 114, Figure 11) and to the block decoder 130 (Figure 19) so as to set up the blocks, in block register 118 (Figure 13), bearing in mind that the block can arise within stage S or may already be stored in the register from any of the stages from earlier set-up phases (if not reset). The results of arbitration create signals S0 and S1 which are latched (centre, Figure 15) as SEL 0 and SEL 1. The latter are the signals which actually switch the cell and request signals through crossbar switches 108, 110 respectively. The fairness register 114 is set to show the paths created by the arbitration. The cell is transferred through the crossbar switch 108. Because of the delay created by the request delay 112, the SEL 0 and SEL 1 signals arrive at the crossbar switch 108 two clock cycles late relative to the RQ (request signals) so that the first two bits of the CCF (i.e. NC_s, $FULL_s$) are lost and the cell passes through the switch 108 to one or both outputs with NC_{s-1}, $FULL_{s-1}$ at the head of the CCF. The request signals similarly pass to the same outputs in crossbar switch 110. The cell is transferred to the forcer 120 (Figure 14) where NC_{s-1}, $FULL_{s-1}$ is forced to "1,1" if the CCF for stage S has indicated that there is to be full replication of the cell in each subsequent stage.

The cell and its corresponding request signals then pass to stage S-1 in the multicast fabric where the set-up phase repeats. This continues through the remaining stages of the fabric, at which time the set-up phase finishes. If at any stage the cell becomes blocked, the blocking signal ripples back through the preceding stages (and sets the block registers thereof) to the input port.

The data phase then commences and, if no blocking exists, the cell is clocked out through the multicast fabric to the appropriate output port(s). In such an event, no blocking signal is received at the input port, which will signal that there has been successful transmission through the switch fabric. The input buffer is then cleared ready to receive another incoming cell and the retry counter is set once more.

If blocking arises, this signals to the retry counter at the input port that an unsuccessful transmission took place. The retry counter decrements by 1 and commences the next set-up phase, trying to send the same unsuccessful cell again. As the fairness register 114 can "toggle" the result of arbitration by the arbitrator 126 there is a chance that this retry might be more successful. This does not always follow as the fabric will, by then, possibly be trying to set up paths for new cells that have just arrived at the fabric at this first retry set-up phase. A series of set-up and data phases follow until the cell is successfully transmitted or the retry counter decrements to 0. Once the retry counter has decremented to 0, the RESET signal is generated (reset signal decoder, Figure 20) which ripples forward through the fabric on the next set-up phase to clear the block registers for the path for the unsuccessful cell.

CLAIMS

1. A multicast switch element suitable for connection to a plurality of similar such elements in a matrix to form a multicast switch fabric, said element having a plurality of inputs and a plurality of outputs and a switch therebetween for passing through to the outputs cells of information received at the inputs and for replicating copies of the cell information in accordance with instructions associated with the cell information, and means associated with the switch for decoding said instructions in two sections: one section determining the number of times the cell information is to be replicated within the switch, and the other section determining whether or not the cell is to be replicated fully in any subsequent switch elements to which the element may be connected in the multicast switch fabric.

2. A switch element according to claim 1 comprising means for indicating at the input whether the switch will be able successfully to transmit the cell information to the outputs in accordance with said instructions.

3. A switch element according to claim 2 wherein said indicating means comprises a block register for storing information as to whether a path in the switch is blocked or that a path leading from an output of the switch element to any subsequent switch elements, to which the element may be connected in the multicast switch fabric, is blocked.

4. A switch element according to any of claims 1 to 3 wherein said decoding means comprises means for decoding header information from the cell to obtain said decoded instructions.

5. A switch element according to claim 4 which comprises means for removing the header information relevant to the element from the cell so that the cell is transmitted from the outputs without such removed header information.

6. A switch element according to claim 5 wherein said removing means comprises means for delaying transmission of the cell through the switch so that said header information is lost.

7. A switch element according to any of claims 1 to 6 which additionally comprises means for overwriting information into the instructions associated with the outgoing cell, in the event that the decoding means determines from said other section of information that the cell is to be fully replicated in all subsequent switch elements.

8. A switch element according to any of claims 1 to 7 which additionally comprises a fairness register for storing successful requests for outputs in one cycle of operation of the switch element and an arbitration means for determining, in the event of a collision of requests for an output from more than one input, which input would be granted the output request in the next cycle of operation of the switch, based upon the information stored in the fairness register.

9. A switch element according to any of claims 1 to 8 when in the form of a 2 x 2 switch element.

10. A multicast switch fabric having an M x M matrix of connected switch elements as claimed in any of claims 1 to 9, an input port and an output port associated with each input and output respectively of the fabric, each said input port having an input buffer for storing a cell for transmission through the fabric, and means for retaining any such cell in said buffer until cell transmission is fully successful, or a maximum number of retries at transmission has been reached.

11. A multicast switch fabric having an M x M matrix of connected switch elements, an input port and an output port associated with each input and output respectively of the fabric, each input port having an input buffer for storing a cell for transmission through the fabric, means for providing a set-up phase and a data phase for each cycle of attempted

transmission of cells through the fabric, the set-up phase attempting either successfully or unsuccessfully to set-up desired paths through the fabric for a cell held in the input buffer, and in the event of being unsuccessful to store a block signal at each switch element in the unsuccessful path and to transmit the block signal back to the input port, the data phase transmitting the cell through the fabric if a successful, unblocked path is established, and means to repeat the set-up and data phases for the cell in the event of a block signal being received by the input port.

12. A multicast switch fabric comprising an M x M matrix of connected switch elements, an input port and an output port associated with each input and output respectively of the fabric, each switch element comprising a block register to store a signal indicating that a desired path through the fabric including that element will be unsuccessful, means for transmitting block signals back to the relevant input port, and means for attempting to retransmit the cell through the desired path until the block signal is removed and the cell is transmitted completely through the fabric successfully, or a maximum number of retries at transmission is reached.

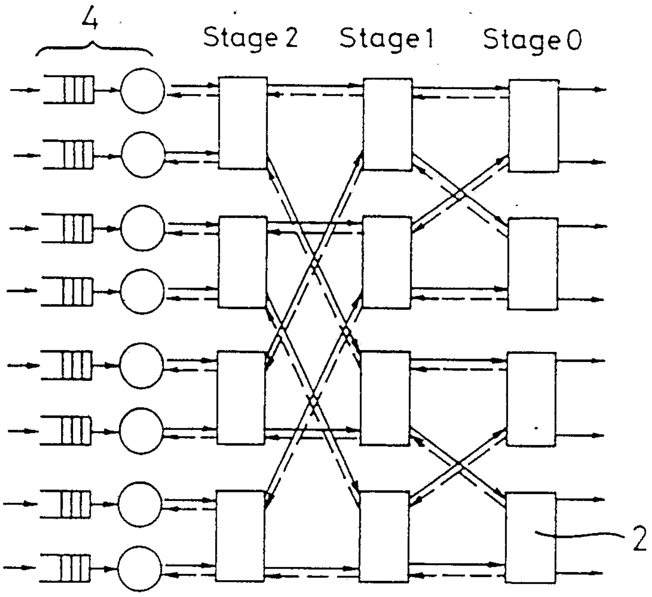


FIG. 1

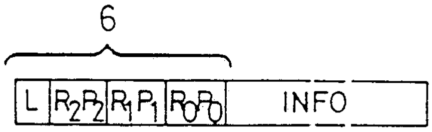


FIG. 2

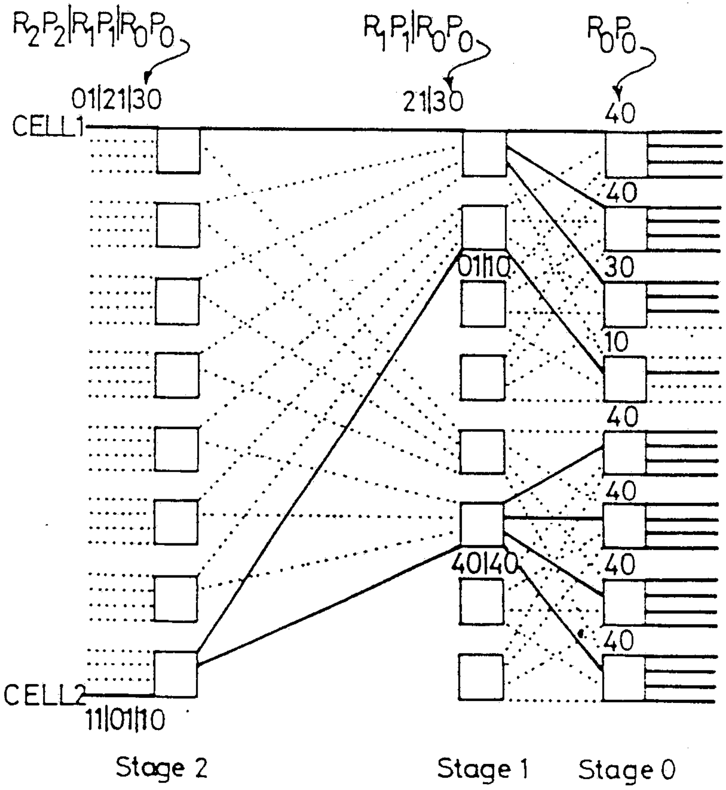


FIG. 3

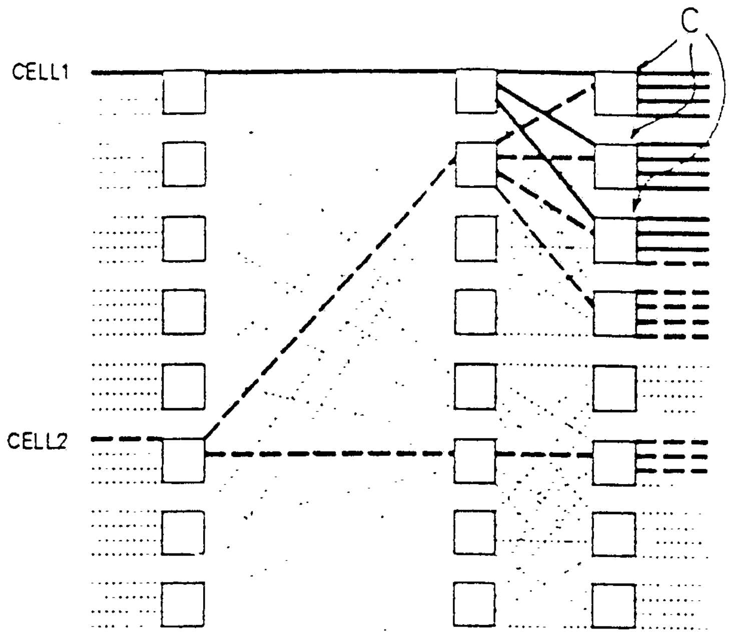


FIG. 4

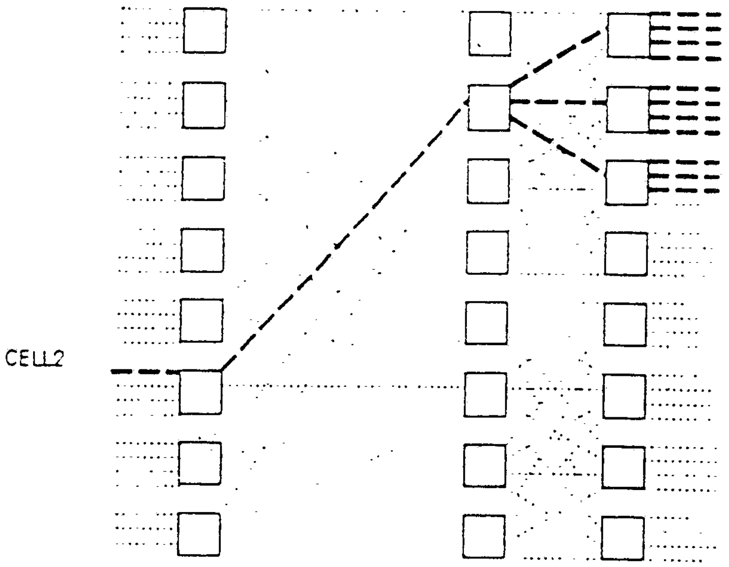


FIG. 5

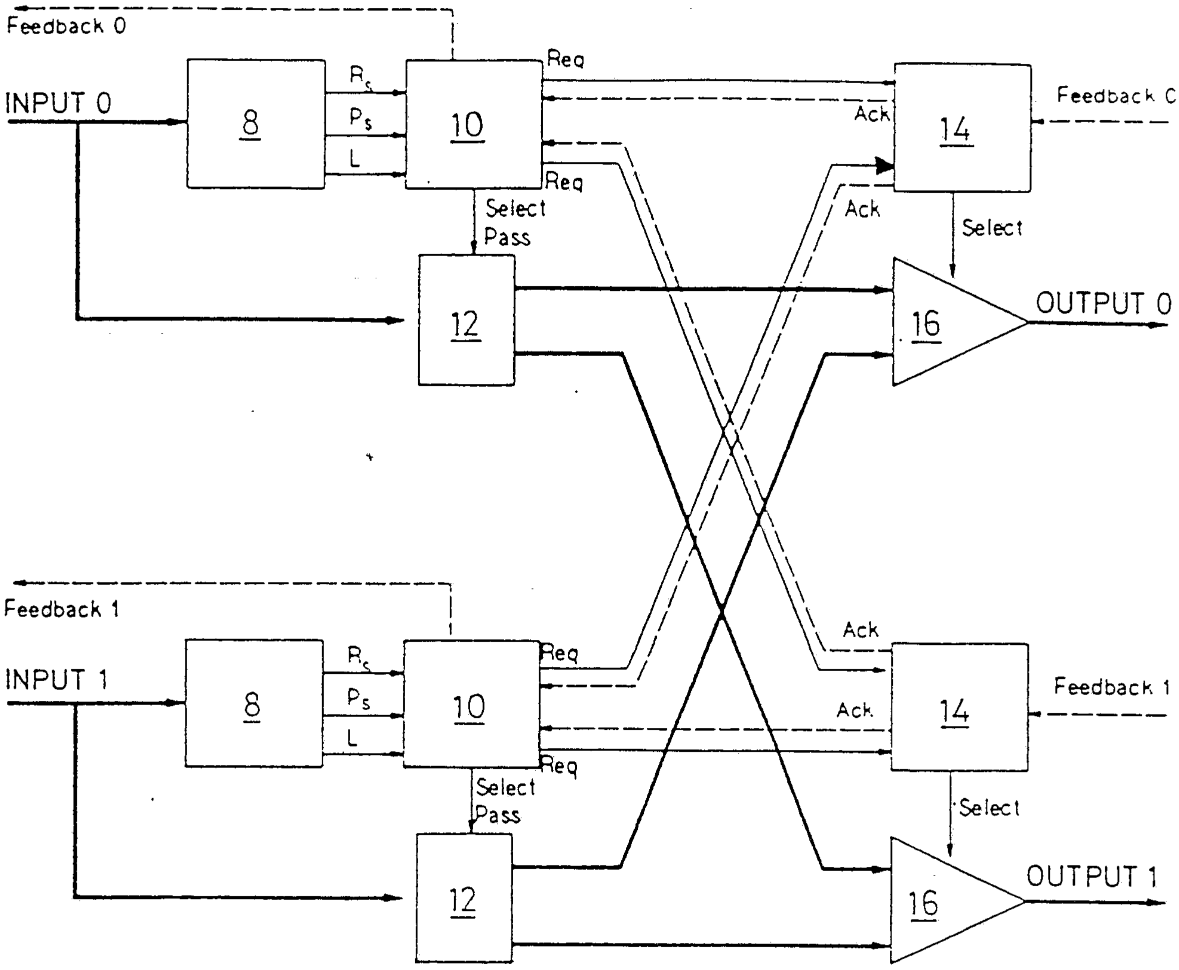

Feedback 0
INPUT 0
8
R_c
P_s
L
10
Req
Ack
Req
Select
Pass
12
14
Feedback C
Ack
Select
16
OUTPUT 0
Feedback 1
INPUT 1
8
R_c
P_s
L
10
Req
Req
Select
Pass
12
Ack
Ack
14
Feedback 1
Select
16
OUTPUT 1

FIG. 6

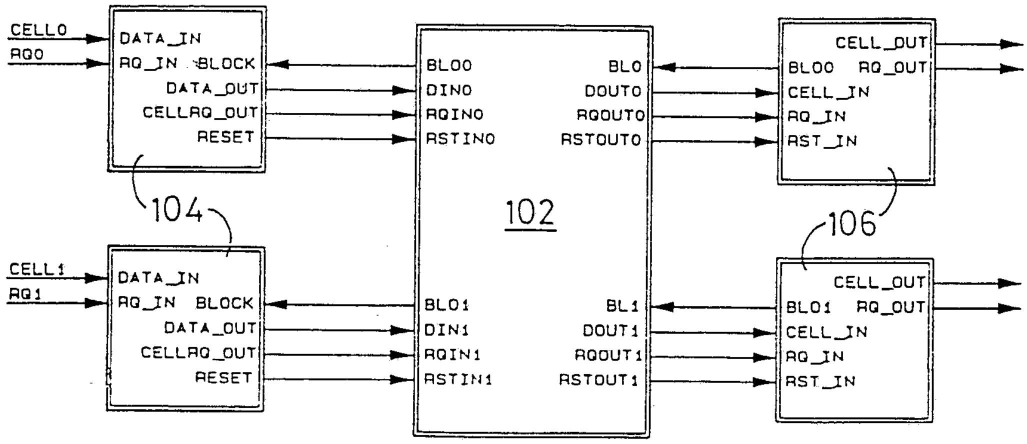

CELL0
RQ0
DATA_IN
RQ_IN BLOCK
DATA_OUT
CELLRQ_OUT
RESET
104
CELL1
RQ1
DATA_IN
RQ_IN BLOCK
DATA_OUT
CELLRQ_OUT
RESET
BLO0
DIN0
RQIN0
RSTIN0
BL0
DOUT0
RQOUT0
RSTOUT0
102
BLO1
DIN1
RQIN1
RSTIN1
BL1
DOUT1
RQOUT1
RSTOUT1
CELL_OUT
BLO0 RQ_OUT
CELL_IN
RQ_IN
RST_IN
106
CELL_OUT
BLO1 RQ_OUT
CELL_IN
RQ_IN
RST_IN

FIG. 7

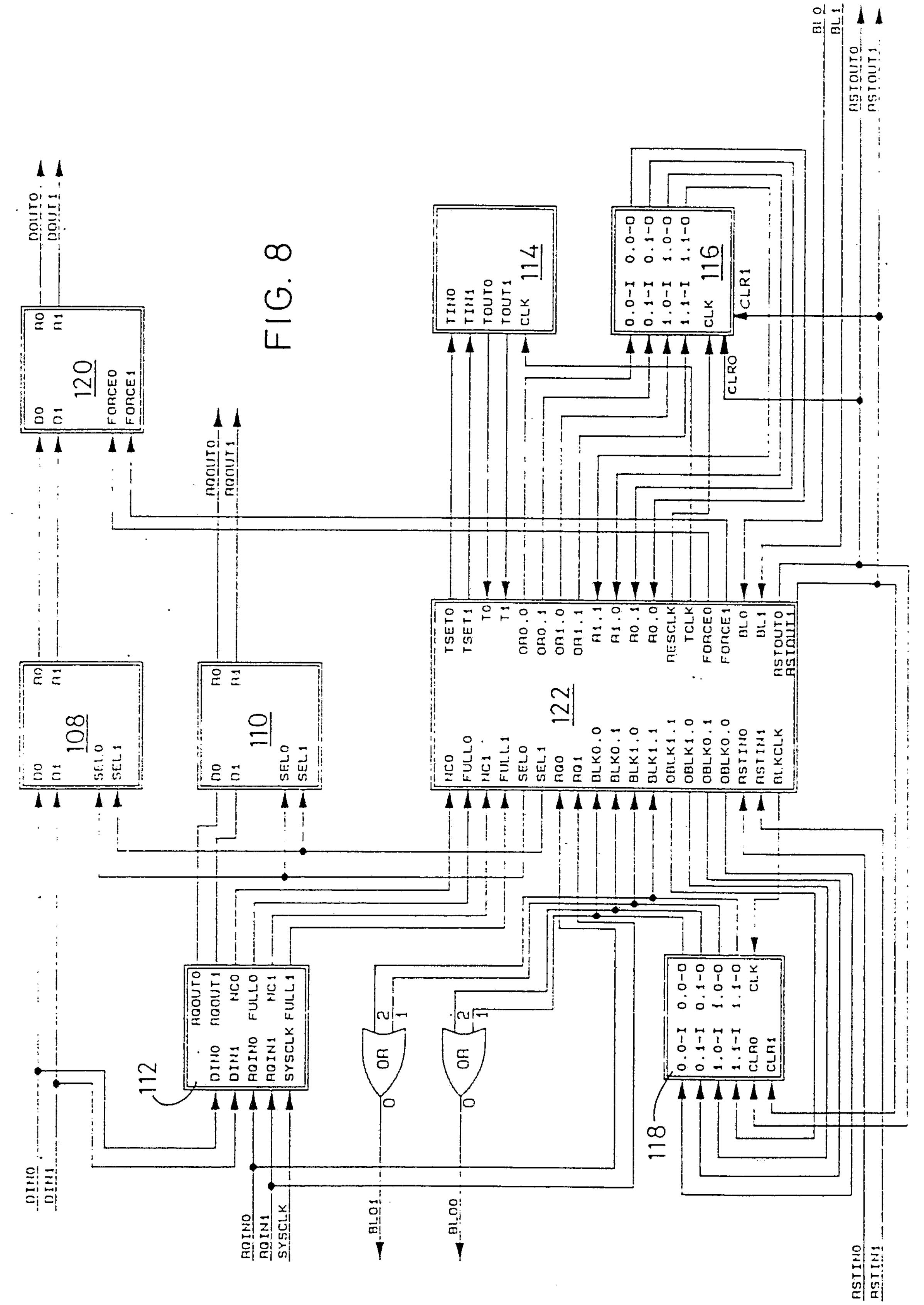

DOUT0
DOUT1
120
D0
D1
FORCE0
FORCE1
R0
R1
108
SEL0
SEL1
110
RQOUT0
RQOUT1
112
DIN0
DIN1
RQIN0
RQIN1
SYSCLK
NC0
FULL0
NC1
FULL1
OR
OR
2
1
0
BL01
BL00
114
TIN0
TIN1
TOUT0
TOUT1
CLK
116
0.0-I
0.1-I
1.0-I
1.1-I
0.0-0
0.1-0
1.0-0
1.1-0
CLR0
CLR1
122
TSET0
TSET1
T0
T1
OR0.0
OR0.1
OR1.0
OR1.1
R1.1
R1.0
R0.1
R0.0
RESCLK
TCLK
FORCE0
FORCE1
BL0
BL1
RSTOUT0
RSTOUT1
SEL0
SEL1
RQ0
RQ1
BLK0.0
BLK0.1
BLK1.0
BLK1.1
OBLK1.1
OBLK1.0
OBLK0.1
OBLK0.0
RSTIN0
RSTIN1
BLKCLK
118
RSTIN0
RSTIN1

FIG. 8

FIG. 9

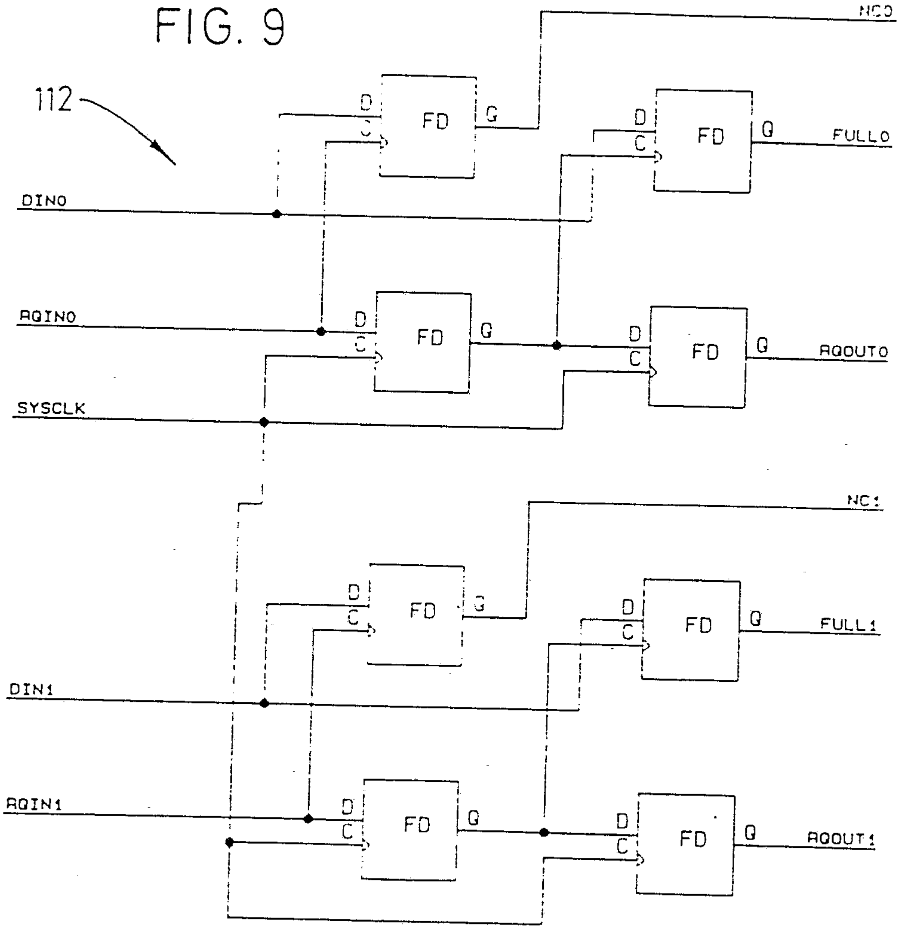
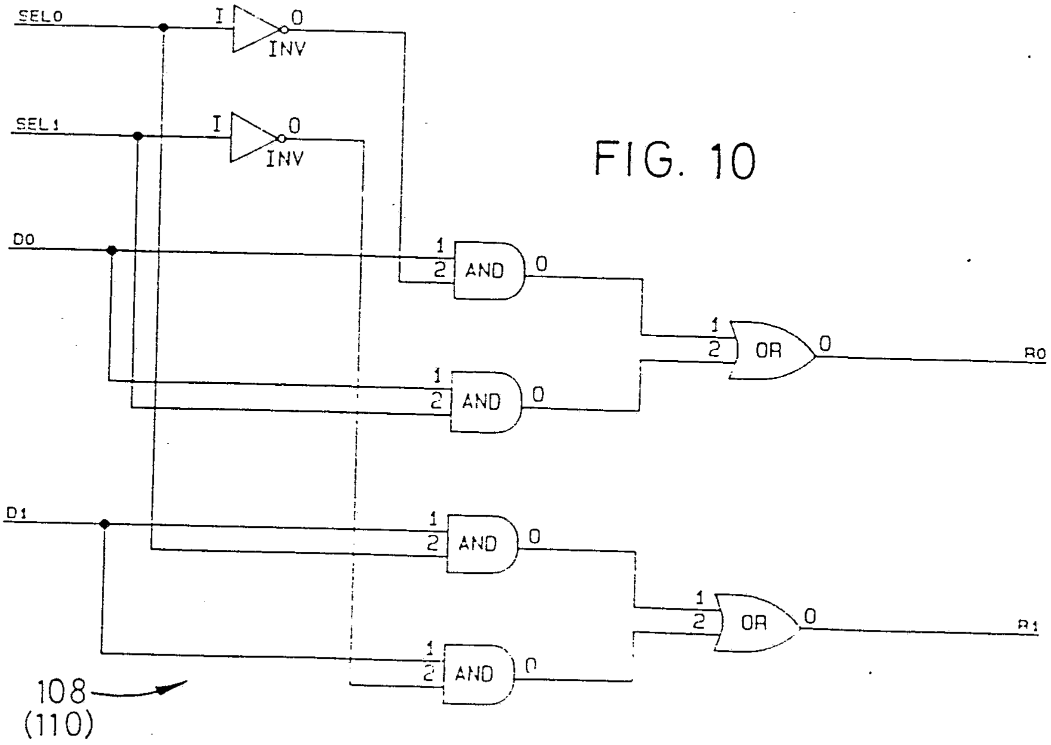

112
NC0
D
C
FD
Q
FULL0
DIN0
RQIN0
RQOUT0
SYSCLK
NC1
FULL1
DIN1
RQIN1
RQOUT1

FIG. 10


SEL0
I
O
INV
SEL1
D0
1
2
AND
0
OR
R0
D1
R1
108
(110)

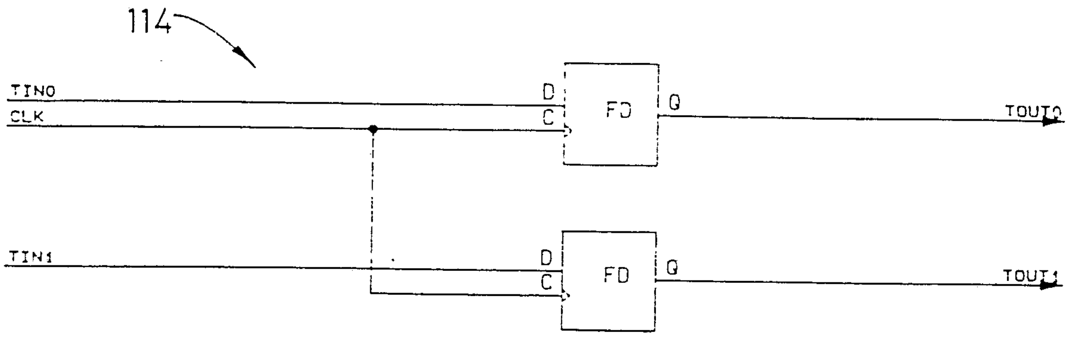

114
TIN0
CLK
D
C
FD
Q
TOUT0
TIN1
D
C
FD
Q
TOUT1

FIG. 11

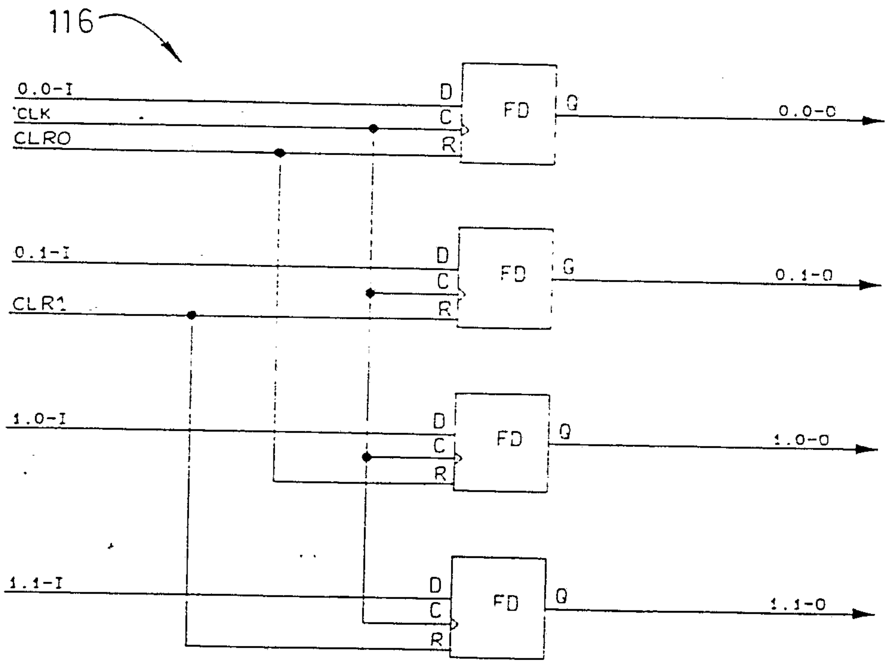

116
0.0-I
CLK
CLR0
D
C
R
FD
Q
0.0-O
0.1-I
CLR1
D
C
R
FD
Q
0.1-O
1.0-I
D
C
R
FD
Q
1.0-O
1.1-I
D
C
R
FD
Q
1.1-O

FIG. 12

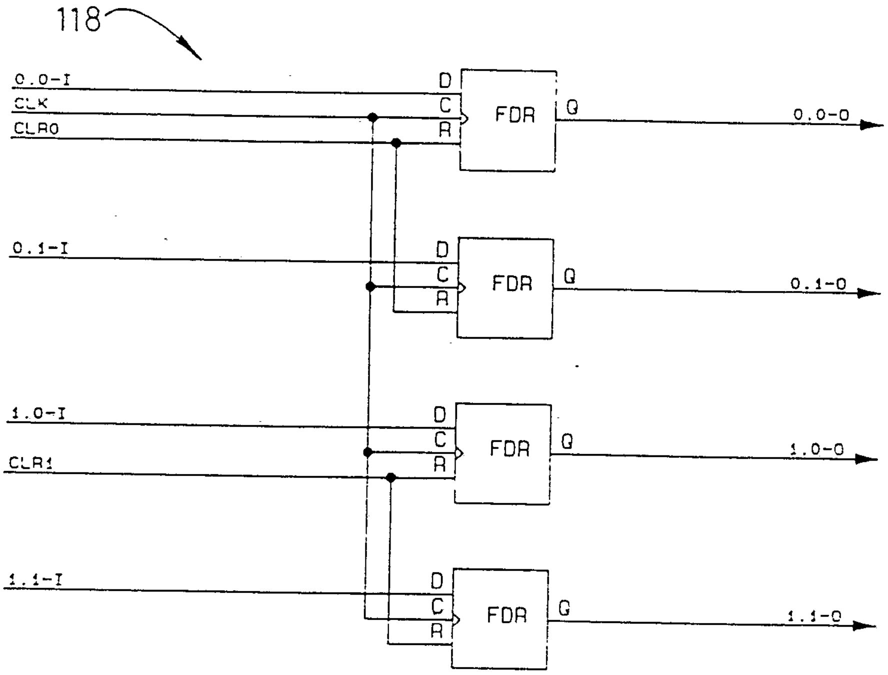

118
0.0-I
CLK
CLR0
D
C
R
FDR
Q
0.0-0
0.1-I
D
C
R
FDR
Q
0.1-0
1.0-I
CLR1
D
C
R
FDR
Q
1.0-0
1.1-I
D
C
R
FDR
Q
1.1-0

FIG. 13

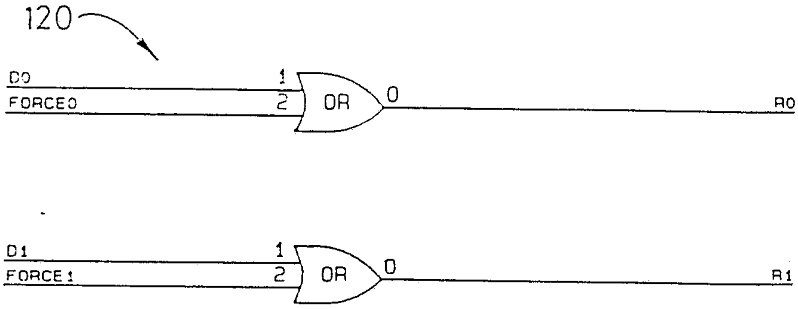

120
D0
FORCE0
1
2
OR
0
R0
D1
FORCE1
1
2
OR
0
R1

FIG. 14

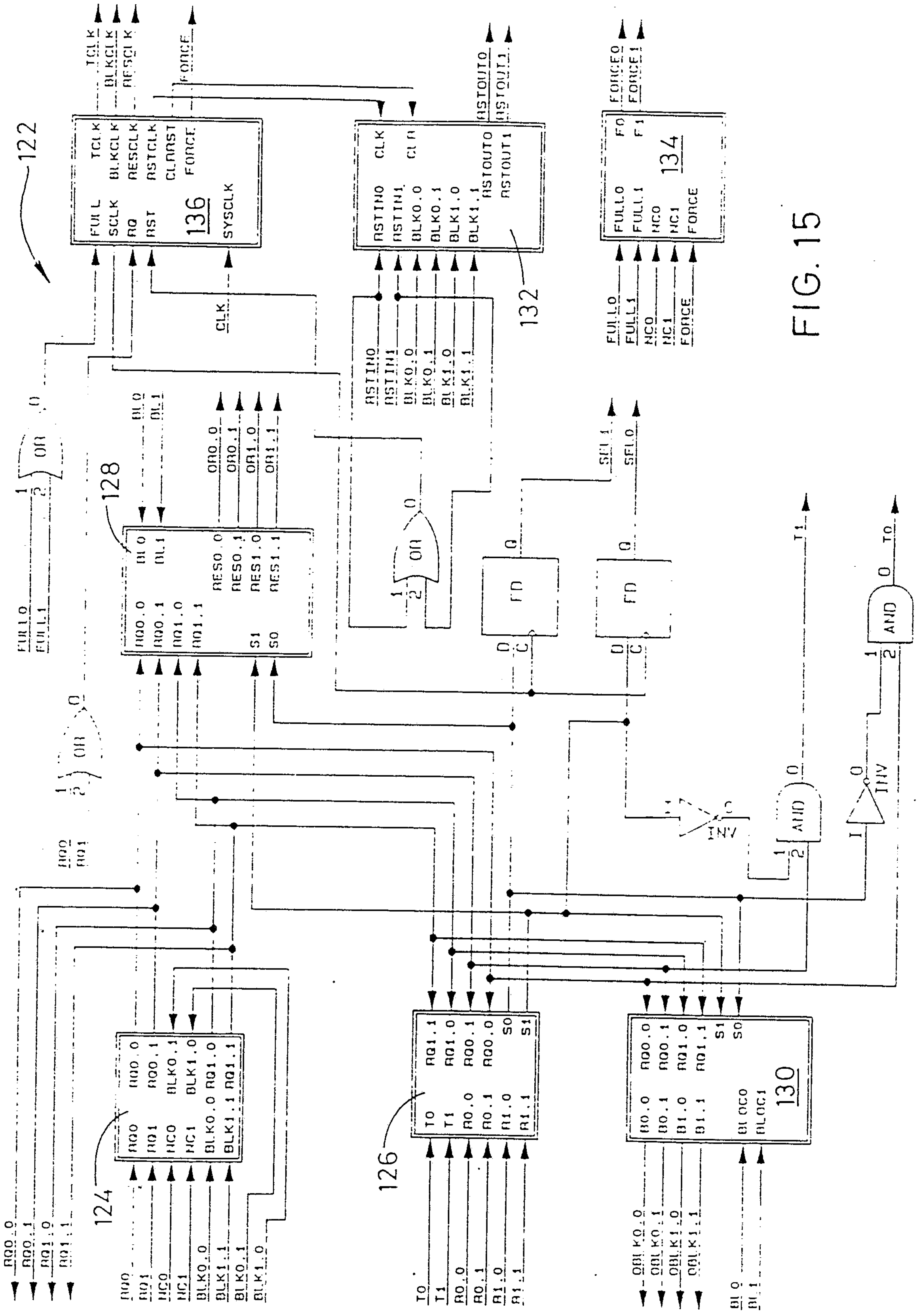

122
136
FULL
SCLK
RQ
RST
SYSCLK
TCLK
BLKCLK
RESCLK
RSTCLK
CLRRST
FORCE
132
RSTIN0
RSTIN1
BLK0.0
BLK0.1
BLK1.0
BLK1.1
CLK
CLR
RSTOUT0
RSTOUT1
134
FULL0
FULL1
NC0
NC1
F0
F1
FORCE0
FORCE1
128
RQ0.0
RQ0.1
RQ1.0
RQ1.1
BL0
BL1
RES0.0
RES0.1
RES1.0
RES1.1
S1
S0
OR0.0
OR0.1
OR1.0
OR1.1
OR
AND
INV
FD
D
C
Q
SEL1
SEL0
T1
T0
124
RQ0
RQ1
NC0
NC1
BLK0.0
BLK1.1
BLK0.1
BLK1.0
RQ0.0
RQ0.1
RQ1.0
RQ1.1
126
T0
T1
R0.0
R0.1
R1.0
R1.1
130
B0.0
B0.1
B1.0
B1.1
BLOC0
BLOC1
OBLK0.0
OBLK0.1
OBLK1.0
OBLK1.1

FIG. 15

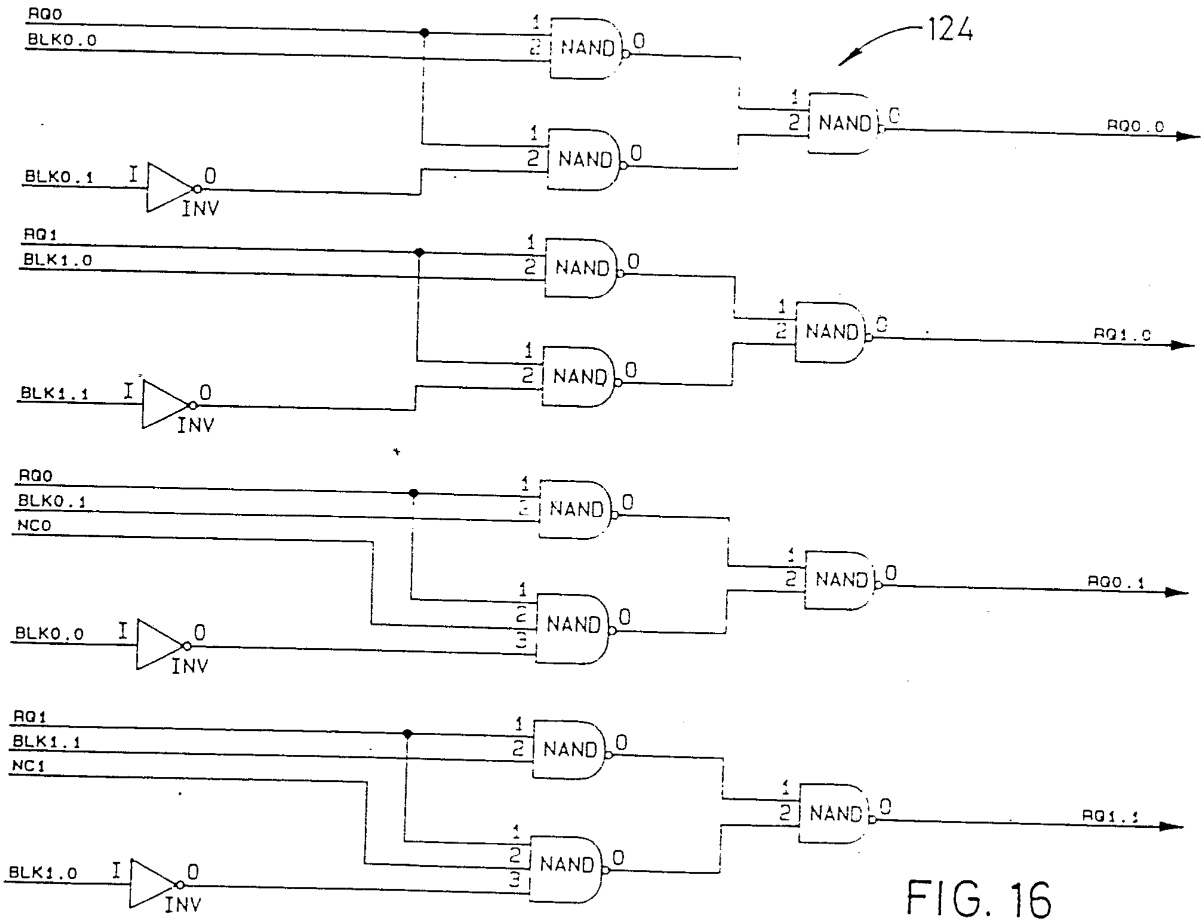

124
RQ0
BLK0.0
BLK0.1
RQ1
BLK1.0
BLK1.1
RQ0
BLK0.1
NC0
BLK0.0
RQ1
BLK1.1
NC1
BLK1.0
NAND
INV
I
0
1
2
3
RQ0.0
RQ1.0
RQ0.1
RQ1.1

FIG. 16

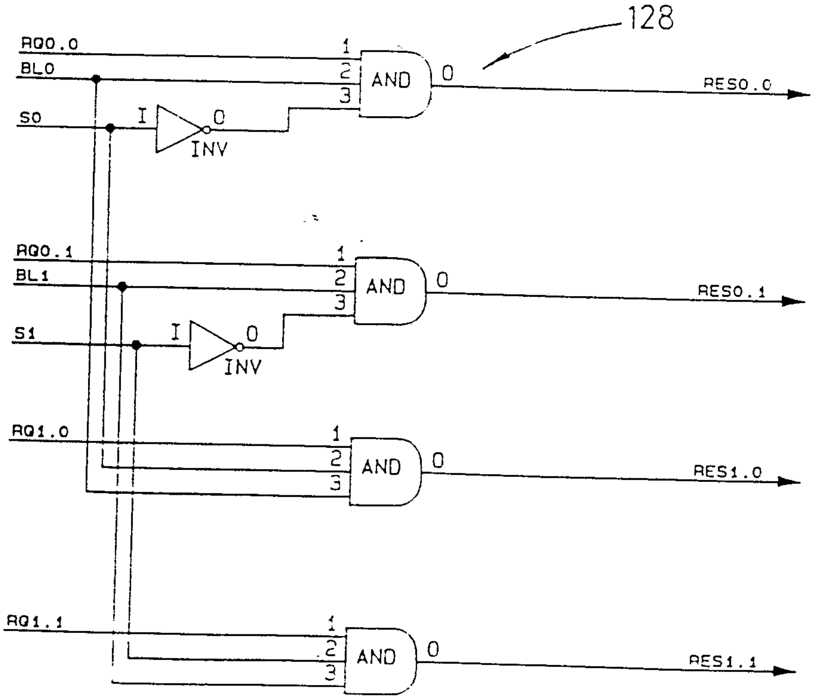

128
RQ0.0
BL0
S0
RQ0.1
BL1
S1
RQ1.0
RQ1.1
AND
INV
I
0
1
2
3
RES0.0
RES0.1
RES1.0
RES1.1

FIG. 18

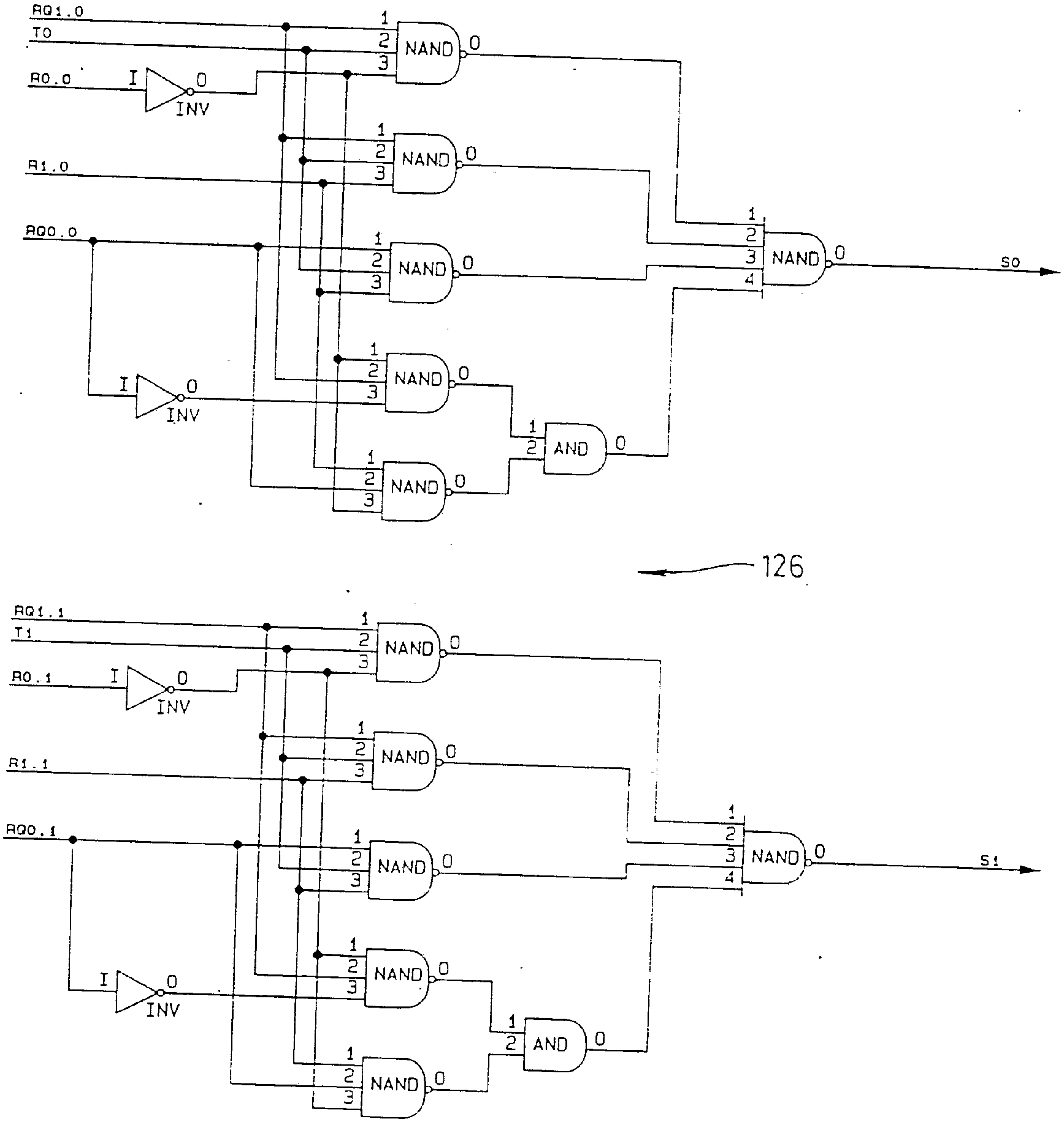

RQ1.0
T0
R0.0
I
0
INV
R1.0
RQ0.0
NAND
AND
1
2
3
4
S0
126
RQ1.1
T1
R0.1
R1.1
RQ0.1
S1

FIG. 17

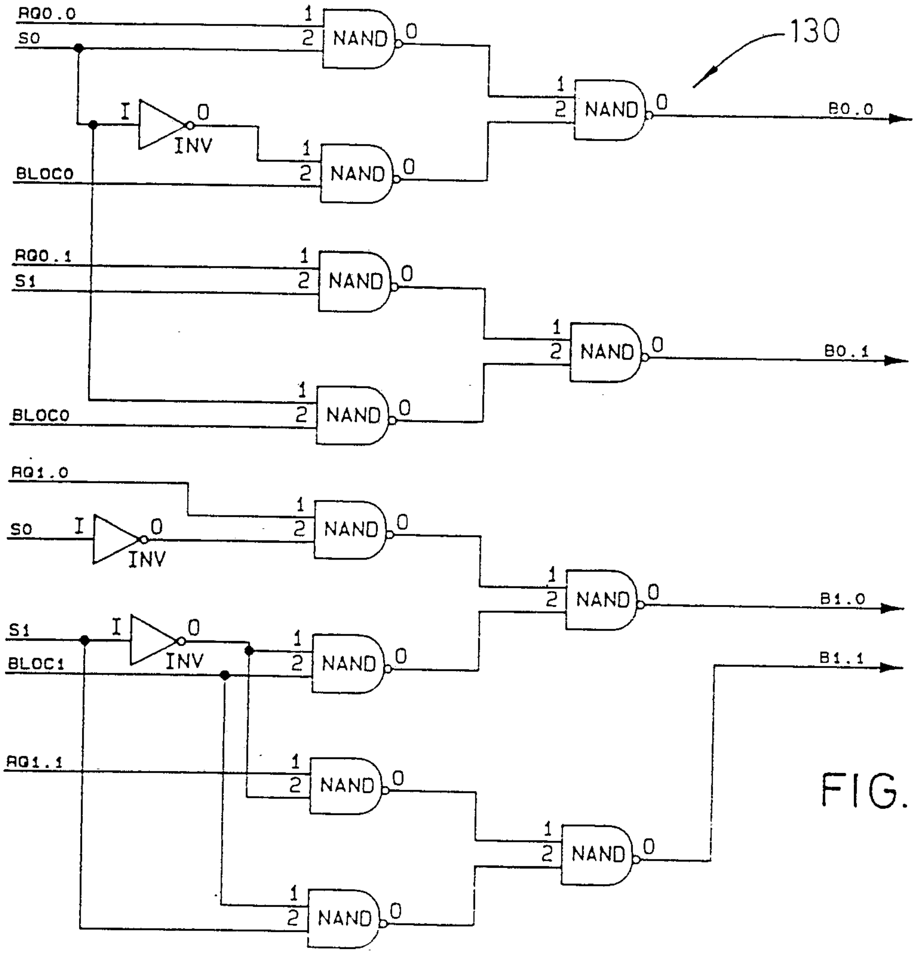
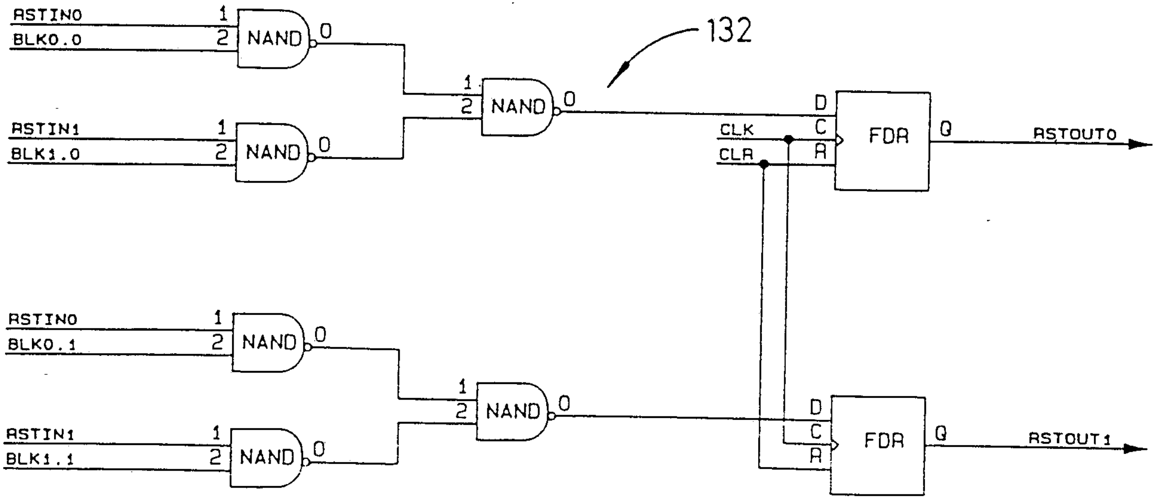

RQ0.0
S0
BLOC0
RQ0.1
S1
BLOC0
RQ1.0
S0
S1
BLOC1
RQ1.1
NAND
INV
I
0
1
2
130
B0.0
B0.1
B1.0
B1.1

FIG. 19

FIG. 20


RSTIN0
BLK0.0
RSTIN1
BLK1.0
RSTIN0
BLK0.1
RSTIN1
BLK1.1
NAND
1
2
0
132
CLK
CLR
D
C
R
FDR
Q
RSTOUT0
RSTOUT1

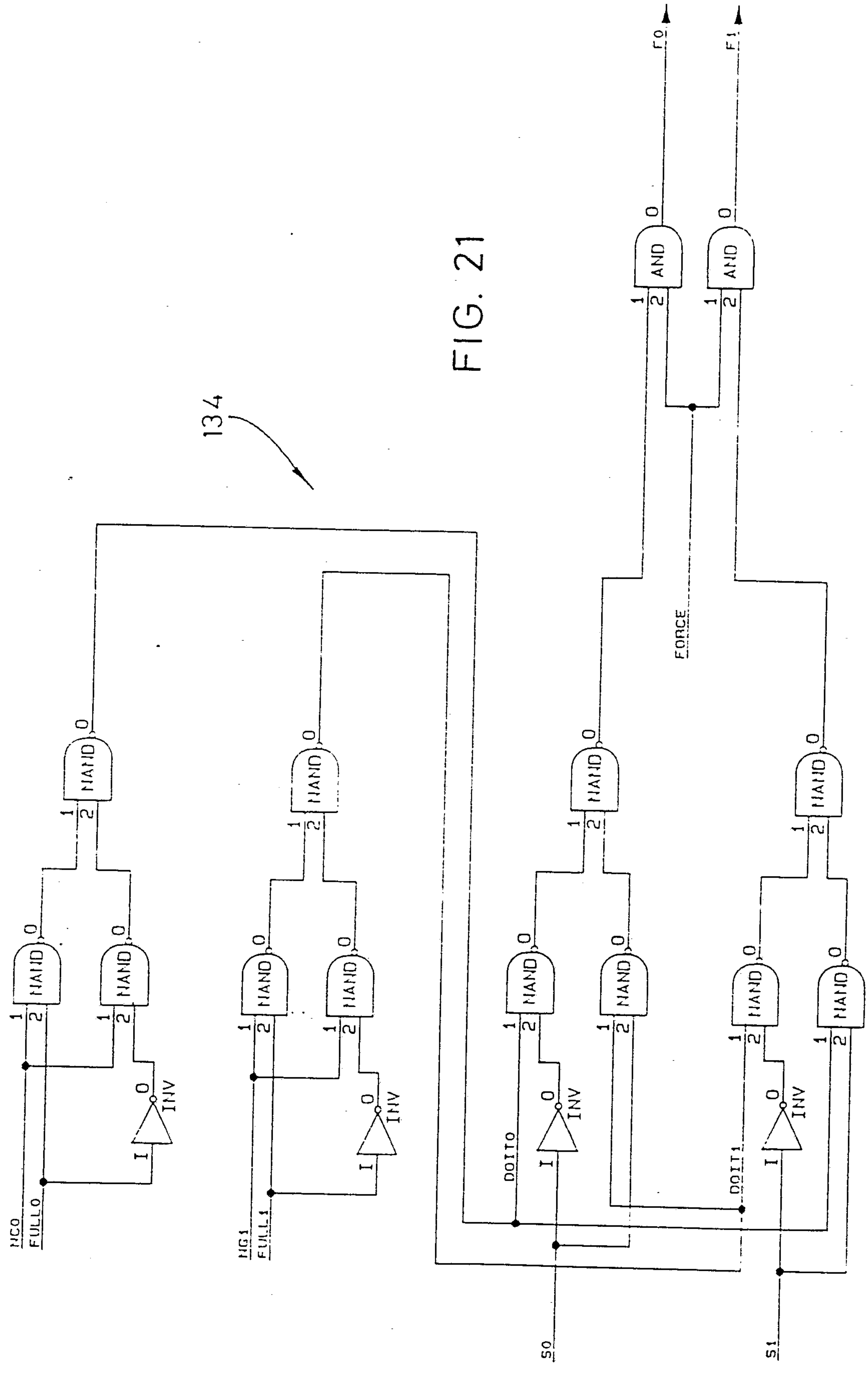

134
F0
F1
AND
NAND
INV
FORCE
DOIT0
DOIT1
S0
S1
NC0
FULL0
NG1
FULL1

FIG. 21

INTERNATIONAL SEARCH REPORT

International Application No PCT/GB 91/01909

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[6]

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

Int.Cl. 5 H04L12/56; H04L12/18

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
Int.Cl. 5	H04L

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched[8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category °	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No.[13]
A	IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATION. vol. 6, no. 9, December 1988, NEW YORK US pages 1455 - 1467; T.T.LEE: 'Nonblocking Copy Networks for Multicast Packet Switching' cited in the application see page 1456, left column, line 6 - right column, line 16 see page 1459, right column, line 1 - line 14 ---	1,2,4,7

-/--

° Special categories of cited documents : [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

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

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

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

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published 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; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search 03 FEBRUARY 1992	Date of Mailing of this International Search Report 11. 02. 92
International Searching Authority EUROPEAN PATENT OFFICE	Signature of Authorized Officer STAESSEN B.F.

3

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category °	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
A	IEEE TRANSACTIONS ON COMMUNICATIONS. vol. 36, no. 6, June 1988, NEW YORK US pages 734 - 743; J.S. TURNER: 'Design of a Broadcast Packet Switching Network' cited in the application see page 736, right column, line 4 - line 17 see page 736, right column, line 51 - page 737, left column, line 31 see page 737, right column, line 13 - page 739, right column, line 11 ---	1,2,4,5, 10-12
A	INTERNATIONAL SWITCHING SYMPOSIUM 1990 MAY 28 - JUNE 1 1990 STOCKHOLM , SE pages 211 - 216; G.W.R. LUDERER ET AL.: 'The evolution of Space Division Packet Switches' see paragraph 7 ---	3,8, 10-12