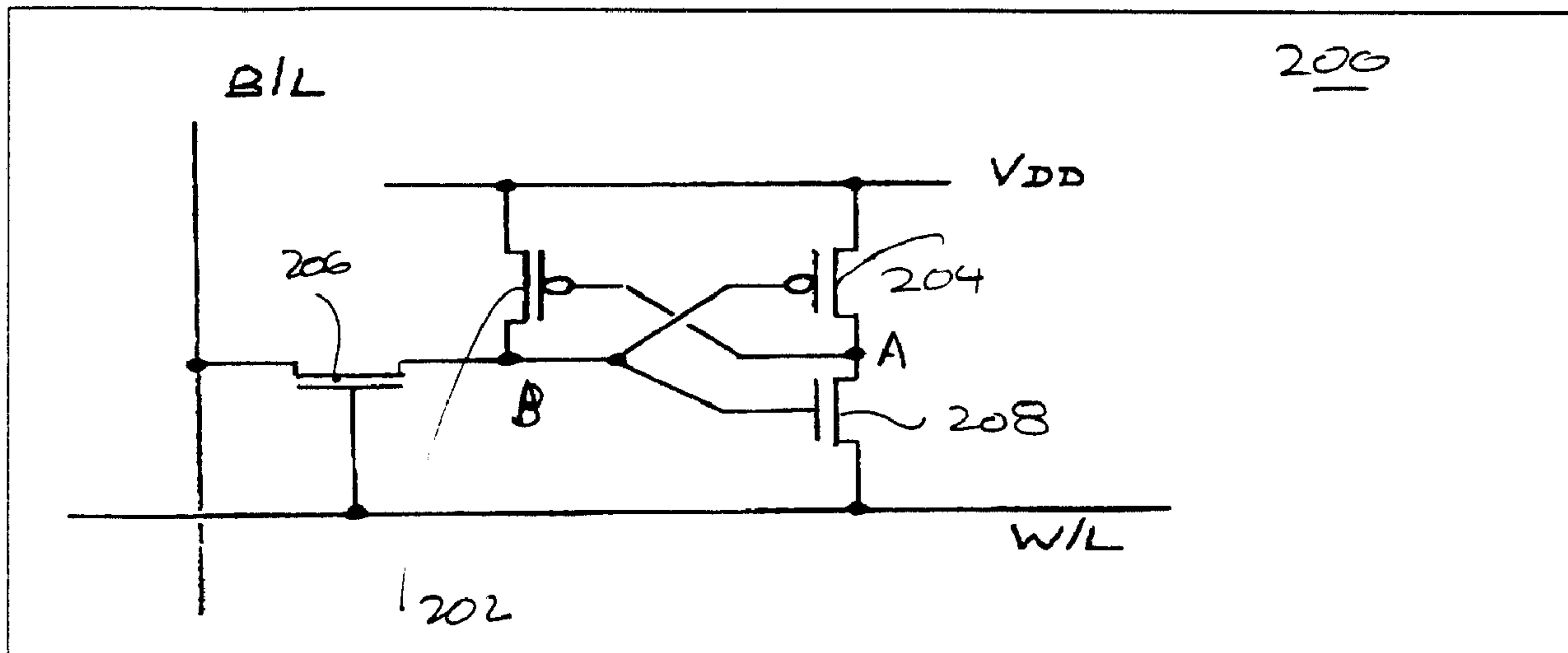




(22) Date de dépôt/Filing Date: 2000/03/03  
(41) Mise à la disp. pub./Open to Public Insp.: 2001/09/03

(51) Cl.Int.<sup>7</sup>/Int.Cl.<sup>7</sup> G11C 11/40  
(71) Demandeur/Applicant:  
MOSAID TECHNOLOGIES INCORPORATED, CA  
(72) Inventeurs/Inventors:  
FOSS, RICHARD C., CA;  
O'CONNELL, CORMAC, CA  
(74) Agent: FASKEN MARTINEAU DUMOULIN LLP

(54) Titre : CELLULE MEMOIRE POUR MEMOIRE IMBRIQUEE  
(54) Title: A MEMORY CELL FOR EMBEDDED MEMORIES



(57) Abrégé/Abstract:

A memory cell comprising an inverting stage, an access transistor coupled between a data line and an input of the inverting stage, the access transistor being responsive to a control signal for selectively coupling the data line and the inverting stage input, a feedback transistor coupled to the inverting stage input and being responsive to an output of the inverting stage for latching the inverting stage in a first logic state and whereby the cell is maintained in a second logic state by a leakage current flowing through the access transistor which is greater than a current flowing through the feedback transistor.

## ABSTRACT

A memory cell comprising an inverting stage, an access transistor coupled between a data line and an input of the inverting stage, the access transistor being responsive to a control signal for selectively coupling the data line and the inverting stage input, a feedback transistor coupled to the inverting stage input and being responsive to an output of the inverting stage for latching the inverting stage in a first logic state and whereby the cell is maintained in a second logic state by a leakage current flowing through the access transistor which is greater than a current flowing through the feedback transistor.

## A MEMORY CELL FOR EMBEDDED MEMORIES

This invention relates to memory devices and more particularly, to a memory cell for embedded memory applications. One particular application discussed herein is constructing  
5 content addressable memories (CAM's) for use in embedded memory systems.

### BACKGROUND OF THE INVENTION

Semiconductor memory has continued to increase in density as a result of a number of technological advances in reducing transistor minimum feature sizes and increased flexibility in  
10 semiconductor device manufacturing capabilities. Both static random access memories (SRAMs) as well as dynamic random access memories (DRAMs) have benefited from advances in commodity as well as embedded implementations. Embedded memory applications typically involve combining memory and other logic functions onto a single semiconductor device resulting in very high bandwidth operation between the memory portion and the other circuitry.  
15 Common applications for embedded memory systems include microprocessor cache memory, microcontroller memory, and various system-on-a-chip applications.

In the networking industry, memory plays an important role in increasing the performance of networking systems in general, and specifically for example in the area of Layer  
3 Fast Ethernet and Gigabit switches. One particular role which memory plays in such switches  
20 is for fast address look-ups. Typically, this type of operation involves comparing an incoming data packet's address information with an existing database consisting of possible addresses indicating where the incoming packet can be forwarded to. This type of operation is very well

suited for implementation using content addressable memory (CAM) especially as network protocols change and databases used for storing such information continue to grow.

Historically CAMs have not gained as widespread usage as DRAMs or SRAMs due to density and fabrication disadvantages. In application specific circuits (ASICs) however, CAMs  
5 have been often used to implement application specific memories for such applications as table look-up and associative computing.

For networking applications, CAM is best suited in applications that require the implementation of high performance wide word search algorithms. In such cases, CAM-based searches provide an advantage over other search algorithms implementations, such as software-  
10 implemented binary tree based searches for example. This is due to the capability of performing searches using very wide words and searching multiple locations in parallel using a CAM.

Typically, data in a CAM is accessed based on contents of its cells rather than on physical locations. A CAM operates by comparing information to be searched, referred to as search data, against the contents of the CAM. When (and if) a match is found, the match address  
15 is returned as the output.

A general background discussion about the various types of CAM cells and their operation is given in the article, "*Content-addressable memory core cells – A survey,*" by Kenneth Schultz in INTEGRATION, the VLSI journal 23 (1997) pg. 171-188. As discussed in the article, CAM cells can be implemented with both SRAM and DRAM type memory cells.  
20 There are clearly advantages and disadvantages to using both types of memories to build CAMs. Generally, DRAM based CAMs have a higher density capacity due to the reduced number of elements required to build a cell as compared with SRAM based CAMs but suffer from the additional complication of requiring periodic refresh in order to maintain the stored data.

Various DRAM based CAM cells have been proposed such as in US Patent No. xxx,xxx,xxx to Mundy, and patent yyy,yyy,yyy to Wade and Sodini and more recently to Lines et al. in US application mm/mmm,mmm assigned to MOSAID Technologies.

In most memory applications, there is an increase in demand for single chip solutions or  
5 so called system-on-a-chip solutions, which require the merging of memory and logic functions on to a single semiconductor chip. DRAM fabrication typically requires special processing steps to construct the cell capacitor structures, such as stacked or trenched cell capacitors. Conversely, SRAM memory cells can be easily implemented by using standard logic processes or so-called “non-DRAM processes”. A disadvantage however of SRAM memory, is that an SRAM cell  
10 typically comprised of 6 transistors or 4 transistors plus 2 resistors, takes up substantially more silicon area than a single transistor plus capacitor found in a typical DRAM cell. When used to construct CAM memory cells, these characteristics of DRAM and SRAM cells are amplified due to the additional complexity required to implement the exclusive NOR function required of a typical CAM cell resulting in relatively large CAM memory cells. And although DRAM based  
15 CAMs provide a density advantage over SRAM based CAMs, the special fabrication process steps typically required for DRAM based technology limit the current potential of DRAM based CAMs in embedded memory applications.

While processes offering DRAM process steps combined with regular logic capability are becoming more available, there is increasing concern that the complexity and cost adders justify  
20 their use only in a limited number of applications. More seriously, the time delay between the availability in the industry of such processes relative to simpler all-logic processes for a given geometry, further impacts the economic case for embedding DRAM. Thus at some die percentage of DRAM to logic, the die will actually be larger if, say, the choice is between SRAM

on an all-logic 0.18 micron process and a merged DRAM/logic 0.25 micron alternative. This is particularly worrying in applications such as Content Addressable Memory (CAM) with a high logic overhead even in stand-alone form which incur an even greater area penalty when embedded.

5           As further considerations, portability between different foundry processes is poorer for the merged process and there are CAD tool inadequacies at this time.

          Accordingly, for embedded memory applications, it is desirable to provide a memory cell which benefits from DRAM based high density characteristics but can be implemented in a pure logic process, requiring no additional fabrication steps for constructing capacitive structures.

10          Preferably, this new cell consists of fewer transistors than typical SRAM memory cells and but does not require a cell capacitor. It is further desirable to use this type of embedded cell to construct embedded CAM cells.

#### **SUMMARY OF THE INVENTION**

15           The present invention seeks to provide a memory cell for embedded memory cell applications having a smaller cell size than conventional SRAM cells, and that is capable of static data storage that is, no refresh of data in the cell is required.

          An advantage of the present invention is that it can replace both regular and embedded SRAM and DRAM cells in embedded memory applications. In particular, it can be built using a  
20          regular logic process with no requirement for additional process steps associated with complex capacitive structures.

          In accordance with this invention, there is provided a memory cell comprising:

(a) an inverting stage;

- (b) an access transistor coupled between a bit line and an input of said inverting stage and being responsive to a control signal received along a control line for selectively coupling said bit line and said inverting stage input; and
- (c) a feedback transistor coupled between said inverting stage input and a power supply and being responsive to an output of said inverting stage for latching said inverting stage in a first logic state and whereby said cell is maintained in a second logic state by a leakage current flowing through said access transistor which is greater than a current flowing through said feedback transistor.

### BRIEF DESCRIPTION OF THE DRAWINGS

These and other features of the preferred embodiments of the invention will become more apparent in the following detailed description in which reference is made to the appended drawings wherein:

**Figure 1** is a schematic diagram of an SRAM cell, according to the prior art;

**Figure 2** is a memory cell, according to an embodiment of the present invention;

**Figure 3** is a ternary CAM cell, according to a further embodiment of the invention;

**Figure 4** is a schematic diagram of an n-channel quad configuration, according to an embodiment of the present invention; and

**Figure 5** is a schematic diagram of a ternary CAM cell, according to a another embodiment of the present invention.

### DESCRIPTION OF THE PREFERRED EMBODIMENTS

Referring to **Figure 1**, there is shown an SRAM cell 100 according to the prior art. The SRAM cell 100 comprises a pair of cross-coupled NMOS devices 102 and 104 for drive

transistors and a pair of PMOS devices 106 and 108 coupled to respective nodes C and D for access transistors. This configuration resembles a loadless CMOS four-transistor (4T) DRAM cell used in the 1970's except that in this prior art implementation, the cross-coupled devices are NMOS and the access devices are PMOS. The access transistors 106 and 108 are connected to  
5 respective bit lines ( $BL / \overline{BL}$ ), with their gates connected to a word line (WL). The circuit arrangement 100 is constructed with the transistors having characteristic such that if the bit lines ( $BL / \overline{BL}$ ) are precharged logic high (VDD) and the manufacturing process is manipulated to ensure that the PMOS devices have higher leakage current than the NMOS devices, the cell 100 will operate similar to a standard resistance loaded SRAM cell. In other words, in order to retain  
10 the data without a refresh cycle, an OFF-state current of the PMOS device  $I_{OFF-P}$  has to be higher than that of the NMOS device  $I_{OFF-N}$ . It may be noted that Intel has used a form of this idea in the early 1970s in a 1K Intel memory device that momentarily pulsed the word line as a "planar refresh".

The cell 100 uses the PMOS access transistors 106 and 108 as load devices to maintain  
15 the data in the cell without the need for a refresh operation. More specifically, in a stand-by case, the bit lines ( $BL / \overline{BL}$ ) are precharged to VDD supply (high) and the word line WL is also precharged to VDD.

Assuming that data in the cell is stored as a high ('H') or VDD level at node C and low ('L') or VSS level at node D, (the opposite state, i.e. 'H' on node D and 'L' on node C could also  
20 be stored, of course) the cell maintains this stored data, i.e. the 'H' at node C by ensuring that the leakage or OFF-state current through the PMOS access transistor 108 is greater than the OFF-state current through the NMOS transistor 104, i.e.  $I_{OFF-P} \gg I_{OFF-N}$ . This is achieved by increasing the threshold voltage of the NMOS device relative to the threshold voltage of the



PMOS device so as to allow more leakage current to flow through the PMOS device 108 than through the NMOS device 104. As a result, the OFF-state current of the NMOS transistor 104 with  $V_{DS}=1.8V$  (i.e. the voltage across its drain-source) is lower than the OFF-state current of the PMOS transistor 108 with  $V_{DS}=0.05V$  by approximately two orders of magnitude.

5           There are a number of ways for biasing the "leakage race" in the circuit of Figure 1. Simply controlling threshold voltages  $V_T$ 's by selective gate ion implantation is one. P-channel devices built with N-type polysilicon gates tend to leak due to adverse work function effects creating a buried channel. If the issue is predominantly one of sub-threshold leakage, then adjusting width and/or lengths may accomplish the desired relative difference in threshold  
10 voltages. Another solution is to apply a bias voltage to the tub or substrate (depending on device polarity) to adjust the threshold voltage. For example, for a circuit having PMOS cross-couples and NMOS access devices and aiming to have NMOS leakage current greater than the PMOS leakage current, by applying a voltage to (or pumping) the n-tub (i.e. the tub in which the PMOS devices reside) to a voltage higher than VDD, would increase the PMOS devices'  
15 threshold voltage,  $V_{TP}$ , thereby lowering the P-sub-threshold current. This approach would only be effective if leakage from the source or drain of the PMOS transistor to the n-type tub did not excessively increase and thereby eliminate the gain in  $V_{TP}$ .

Referring now to **Figure 2** there is shown generally by numeral 200 an improved memory cell, according to an embodiment of the present invention. The cell 200 is comprised of  
20 a pair of crossed-coupled PMOS transistors 202 and 204 each having their respective sources coupled to a VDD supply and transistor 204 having its drain coupled at node A to an NMOS pull-down transistor 208, thereby forming an inverting stage, and transistor 202 having its source coupled at node B to an NMOS access transistor 206 which couples node B to a bit line BL

while its gate is coupled to a word line WL. The NMOS transistor 208 has its source coupled to the word line WL while its gate is coupled to node A of the cross-coupled pair.

Thus, it may be seen that the cell 200 of Figure 2 is an improvement over the cell 100 shown schematically in Figure 1, in that there is a reduction of one bit line and the ground line VSS. In addition, the cell 200 provides at least one "hard" node A for a possible search interrogation in a CAM operation. Just as regular 4T SRAM cell can operate dynamically or statically, if a keeper current is supplied by a resistive load or other similar means, the cell 200 may be operated in either dynamic or static mode. Schematically, the cell 200 is shown "upside down" compared to the cell 100 of Figure 1, in that the transistor types of the cross-coupled and access devices are reversed. However, depending on how the leakages between P and N are controlled, the cell 200 can equally well be implemented with N-channel devices as the cross-coupled transistors 202 and 204 and P-channel transistors as access devices 206 and 208 while the bit line is held normally high.

Referring back to Figure 2, the cell configuration 200 has its word line (WL) normally low and is only pulsed high briefly to turn on the access device 206. Therefore, the word line (WL) serves as the ground for the inverting stage of the circuit. This has a drawback of increasing the word line capacitance, but has the advantage of eliminating a line and simplifying writing to the cell. The operation of the cell may be explained as follows: in a write operation, the bit line BL is set high or low depending on the logic of the data to be written. The word line WL is then pulsed for a predetermined period. During the time period when the WL is high, data is passed to the "soft" node B via the access device 206. Once WL returns to low, data is stored on the "soft" node B and the inverted or "hard" node A.

There are two cases where a change of state of the cell 200 is possible. The first is writing of a low into a cell, which currently stores a high, and the second is the writing of a high into a cell, which currently stores a low. The other two possibilities are the writing of a low into a cell which already stores a low and a writing of a high into a cell which already stores a high.

5 These latter two possibilities are also supported but will not be described in detail since no change of state in the cell occurs.

Firstly, consider the case of writing a low into a cell, which stores a high. Prior to the write operation beginning, the "soft node" B is high and the inverted or "hard node" A is low. As previously mentioned, the write line WL is kept low during standby. Next, a low is loaded  
10 onto the bit line BL and the WL begins to rise turning on the access device 206, thereby bringing node B to a low, which in turn, sets node A high through the pull-up PMOS transistor 204 in the inverting stage. The PMOS transistor 202, which is connected to the "soft" node B, is therefore turned off. Once the WL is turned off, the access device 206 turns off so both the access device 206 and the pull-up PMOS transistor 202 are both off. The low data on node B and high data on  
15 node A are kept in this state through leakage current only. Specifically, the leakage or OFF-state current flowing through the NMOS access device 206 is greater than the OFF-state current flowing through the PMOS pull-up transistor 202 connected to node B. This can be accomplished for example, by setting a threshold voltage of the NMOS transistor 206 to be lower than that of the PMOS transistor 202. This lets the access NMOS device 206 conduct more than  
20 the PMOS device 202. Alternately, the threshold voltages of the NMOS and PMOS devices can be altered by applying a negative back bias voltage to the P-type well in which the NMOS device lies or by applying a slightly lower than VDD voltage to the N-type tub in which the PMOS device lies.

Secondly, consider the case of writing a high into a cell that stores a low. In essence a similar process as described above is executed. A high is placed on the bitline BL. The word line WL is brought high turning ON the access NMOS device 206. This passes a  $V_{DD} - V_{TN}$  level to node B (the threshold voltage drop  $V_{TN}$  occurs across the NMOS access device 206).

5 The NMOS control device 208 is now ready to turn ON. The VDD level at the access transistor 206 occurs when the bit line BL is high. When the WL is brought low again, the NMOS device 208 turns fully ON and pulls node A to a low level. This turns on PMOS device 202 fully thus latching node B high. In this state, no leakage current is needed to keep the data stored on nodes B and A, since the low on node A ensures that node B is maintained high.

10 The following describes the reading operation. For the case of reading a high stored on node B (and a low on node A), the bit line BL begins precharged low and the word line WL level rises. A pulse of current begins to flow into the bit line as the high stored on node B is read onto the bit line with a threshold voltage drop across access transistor 206 so that the voltage on the bit line eventually reaches  $V_{DD} - V_{TN}$ . This voltage difference on the bit line can be detected  
15 using well known DRAM type sensing, by comparing another half bit-line to which is attached a half size dummy cell. Since the data sensed on the bit line BL has to be restored onto node B in order for the cell to retain the correct data, once the data is sensed and amplified on the bit line this high value is written back into node B while the word line remains high. The write back is then completed as the word line falls.

20 For the case of reading a low stored on node B (and a high on node A), the bit line begins precharged and the word line rises. Since there is no voltage difference between the bit line which is precharged low and the value stored on node B, no current flows and the value on node B remains unchanged. Once the word line falls, the value on node B is maintained as described

earlier by ensuring that the leakage current through the NMOS access device 206 is greater than the leakage current through the PMOS feedback transistor 202, i.e.  $I_{OFF-N} \gg I_{OFF-P}$ . This can be accomplished as described earlier by applying a voltage lower than VSS to the p-well of the NMOS device 206 (for example, an on-chip generated voltage supply VBB). This will

5 effectively lower the threshold voltage of NMOS device 206 relative to the threshold voltage of PMOS device 202 and ensure that a larger leakage current flows through NMOS device 206 than through PMOS device 202, thereby maintaining the low value on node B.

It must be remembered that this sense-restore is only needed in a CAM if (a) the cell is operating as a dynamic CAM (DCAM) cell or (b) as a dynamic back-up mode to static mode

10 operation or (c) there is a need to read the cell contents (such as in testing or operating mode). In general however, for common search and compare operations generally performed by CAMs, the read operation is not needed.

In constructing cell 200 of the present invention, use is made of both P and N devices thus a trench isolated process with tight P+ to N+ spacing would be preferable.

15 In a further embodiment of the present invention, the cell configuration of the present invention may be utilized to implement a ternary CAM cell. Any ternary CAM cell should be capable of both storing a "don't care" state and searching with a masked bit. Accordingly, the CAM cell must have three states for each, which in practice requires a double binary cell, i.e. the cell must be able to store a logic "0" a logic "1" and a logic "don't care" and must also be able to

20 mask these three values.

Referring to **Figure 3**, a 10 transistor CAM cell is shown constructed according to an embodiment of the invention, which comprises a pair of cells memory cells 200 as described with reference to Figure 2 and additional devices 306 and 308 for implementing an exclusive OR

(XOR) function 304. There are numerous ways of implementing the XOR function given NMOS and/or PMOS devices each implementation having circuit and layout advantages and disadvantages. In Figure 3, AND gating between the source and the gate of P channel devices 306, 308 is shown. Specifically, PMOS transistors 306 and 308 have their respective source-drain circuits connected between respective hard nodes A and A' and a match line  $\overline{MATCH}$ . Their respective gates are connected to complementary search lines  $SEARCH$  and  $\overline{SEARCH}$  applied to the gates of transistors 306 and 308 respectively. The AND gating operates so that  $SEARCH$  and  $\overline{SEARCH}$  compare with the stored data. The  $\overline{MATCH}$  line will only stay low if hard node A and  $SEARCH$  are both low (and hard node A' and  $\overline{SEARCH}$  are high) or if both A and A' are both low or  $SEARCH$  and  $\overline{SEARCH}$  are high (or of course, A and A' are low and  $SEARCH$  and  $\overline{SEARCH}$  are high). All other combinations result in  $\overline{MATCH}$  being pulled up but only as far as  $VDD - V_{TP}$ , where  $V_{TP}$  is increased by the source-tub bias as it is source-following. This requires a match sensing circuit which will detect the difference between current flowing or not flowing into a level between VSS and  $VSS + V_{TP}$ .

The search / match transistors 306,308 may be implemented as N channel devices without risk of disturbing the soft nodes B and B'. However, the sense line now puts current into the word line which is low thus the match detect circuit must respond to current drawn from the voltage of the match line which is between VDD and  $VDD - V_{TN}$ , where  $V_{TN}$  is source-following enhanced. A difficulty with this implementation is that the basic cell may need to have a voltage  $V_{TP}$  greater than  $V_{TN}$ .

Alternatively, a more conventional 4-transistor XOR circuit may be implemented. This circuit however requires two additional transistors compared to the circuit 300 but in general, these transistors require very little additional area. Referring to Figure 4, an N-channel 4-

transistor circuit configuration 400 is shown. The two extra devices are connected as common-gate, common-source to the "hard" node pull-down transistor 404 and shared source-drain to the search devices 406, 408. This still puts the current into the low word lines but places no sensing restrictions on the match line. Replacing all four transistors with p-channels may be the best solution in this polarity. To avoid any coupling to the soft node B, these transistors may be common-gate, common-source with the transistors whose gates are connected to nodes A, A' and B,B' respectively. This would make writing and searching operations nearly independent.

It should be noted that a binary CAM needs only the one 4T-cell stored with a p-channel XOR quad driven by both nodes. Such a cell thus has 8 transistors and 6 lines and is still capable of doing a masked search though not a stored "don't care".

Clearly, there are a number of factors to weigh in choosing the best compromise between circuit simplicity, area, and rugged operation.

Finally, referring to **Figure 5** a complete version of a ternary CAM implementation according to a preferred embodiment of the present invention is shown. The circuit 500 is based on p-channel access transistors T1 and T1', so the presumption is that P leakage is greater than N, or that the cell operates dynamically. The word line WL is normally high and the match line is pulled down by all cells where a mismatch occurs. A double zero stored on the hard nodes A and A' prevents a pull down as does double zero on the search/search bar lines. As there are actually four states, other functions are also possible. A double one on the hard nodes will inhibit any match being detected regardless of search word unless that search word is masked by a double zero.

The layout would likely cluster T1, T1' and T2, T2' in a tub containing the corresponding devices of the inverted cell above. T1 and T1' might share a gate contact. T3, T4 and T5 will

obviously cluster with the match line likely below VSS. VSS may be a common connection to the inverted row of cells below.

As an additional enhancement to the embodiment of Figure 5, a way of ensuring that data is maintained despite a read operation would be to use a simple form of refresh similar to the planar refresh operation used in the aforementioned Intel 1K DRAM. If a selected cell's word line (which is normally held at VDD in this PMOS access implementation) is periodically dropped to  $VDD - V_{TP}$  while its associated bit line is high, this will top up a high level in the cells storing logic high's on their soft nodes B through a current mirror plus sub-threshold action. This top up current will be easily overcome by NMOS pull-down transistors when turned on. This will most likely work best with the symmetric version of the CAM cell as shown in Figure 5 or with an asymmetric version where the P-channel in the half-flop (T2 or T2') has its source at VDD rather than WL since lowering the WL will reduce the drive to the N-channel holding down a zero on the soft node. This refresh would be as transparent as the static operation described by with reference to Figure 1 and multiple word lines could be glitched in this way.

The basic cell configuration 200 could be used for embedded SRAM applications such as cache memory in microprocessors/microcontrollers.

In a CAM context, if static, the cell can be a "write-only" cell with search logic connected to it. Alternatively, it can be read destructively, its state sensed and restored back in, which is how it would be operated dynamically.

Although the invention has been described with reference to certain specific embodiments, various modifications thereof will be apparent to those skilled in the art without departing from the spirit and scope of the invention as outlined in the claims appended hereto.



**THE EMBODIMENTS OF THE INVENTION IN WHICH AN EXCLUSIVE PROPERTY OR PRIVILEGE IS CLAIMED ARE DEFINED AS FOLLOWS:**

1. A memory cell comprising:
  - (a) an inverting stage;
  - (b) an access transistor coupled between a data line and an input of said inverting stage, said access transistor being responsive to a control signal for selectively coupling said data line and said inverting stage input; and
  - (c) a feedback transistor coupled to said inverting stage input and being responsive to an output of said inverting stage for latching said inverting stage in a first logic state and whereby said cell is maintained in a second logic state by a leakage current flowing through said access transistor which is greater than a current flowing through said feedback transistor.
2. A memory as defined in claim 1, said control line being a word line.
3. A memory as defined in claim 1, said data line being a bit line.
4. A memory device as defined in claim 4, said transistors being PMOS devices and said access transistor being an NMOS device.

«(In Stand-by Cycle)»

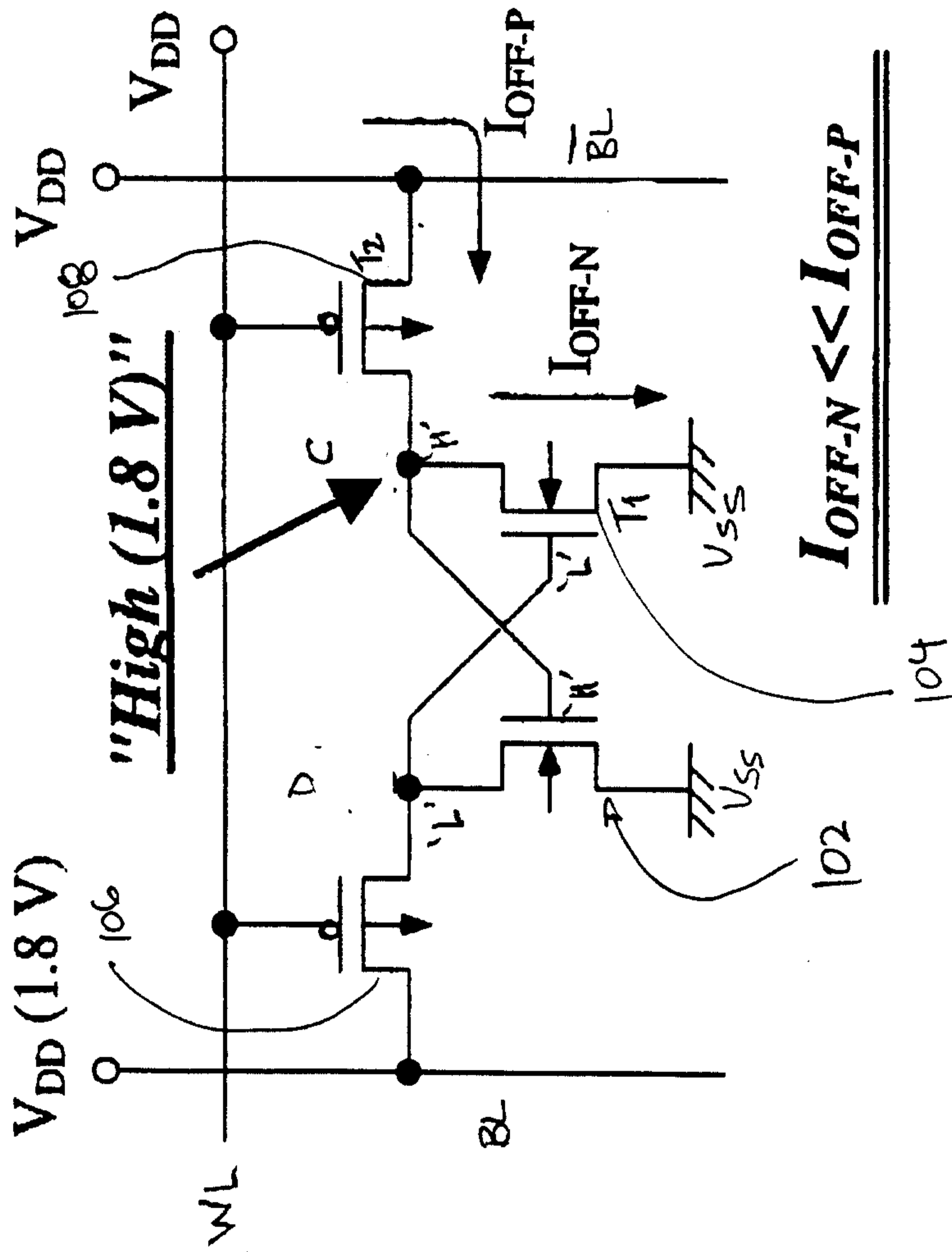
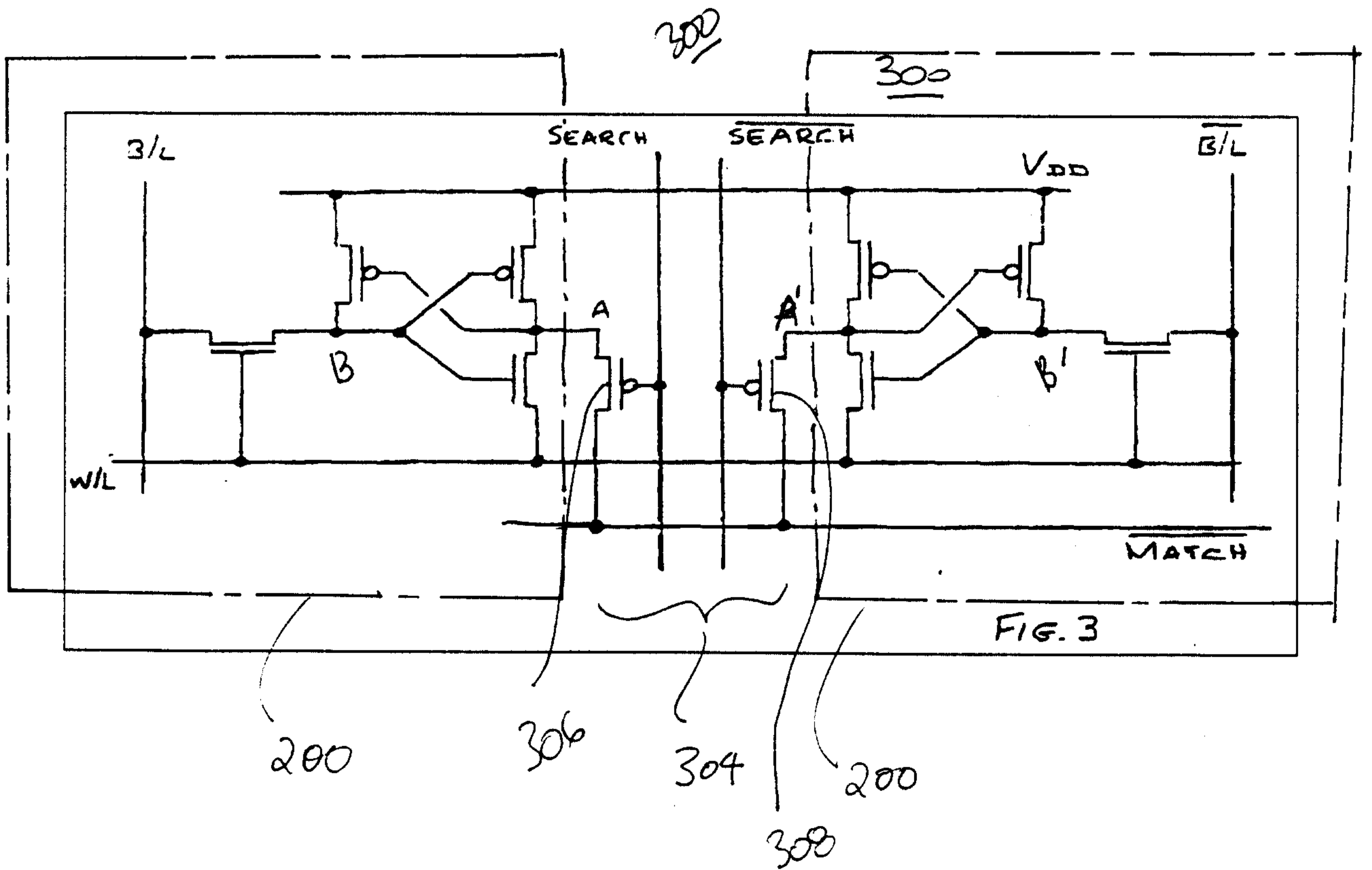
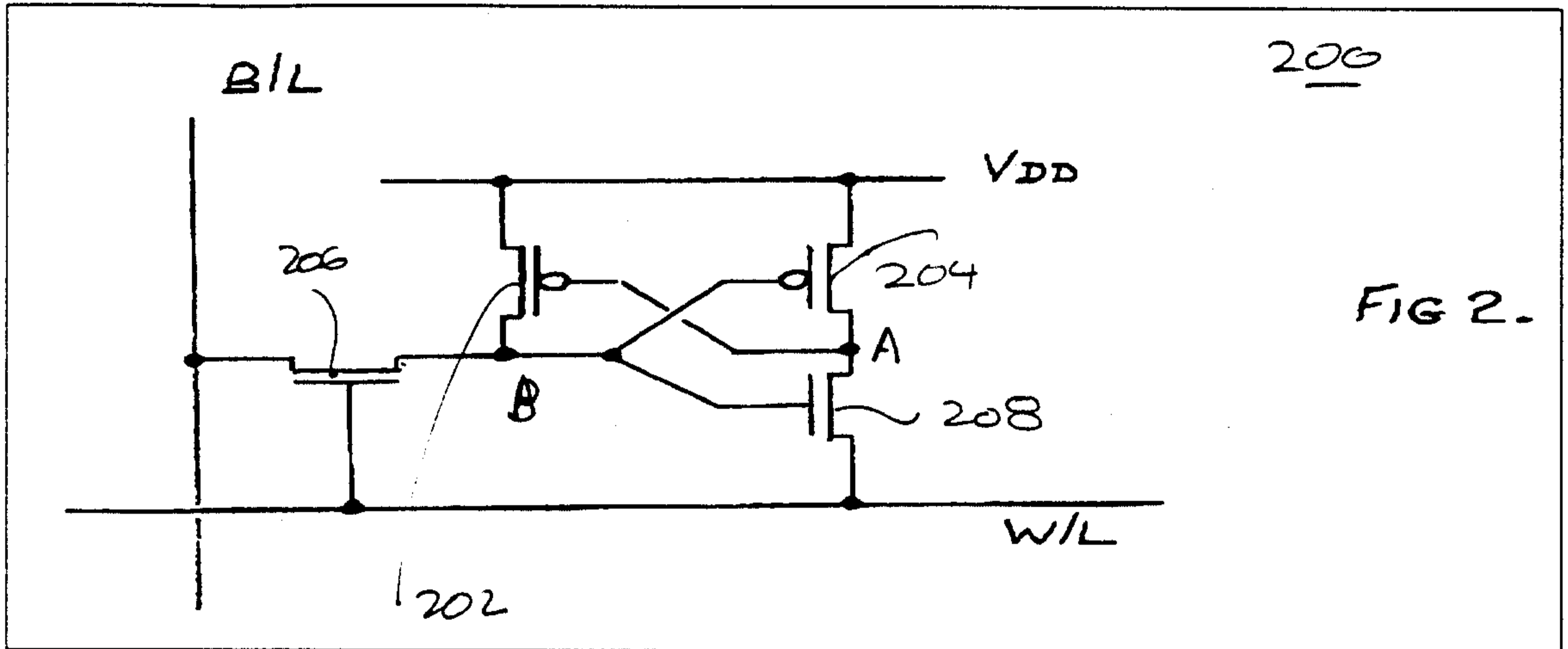
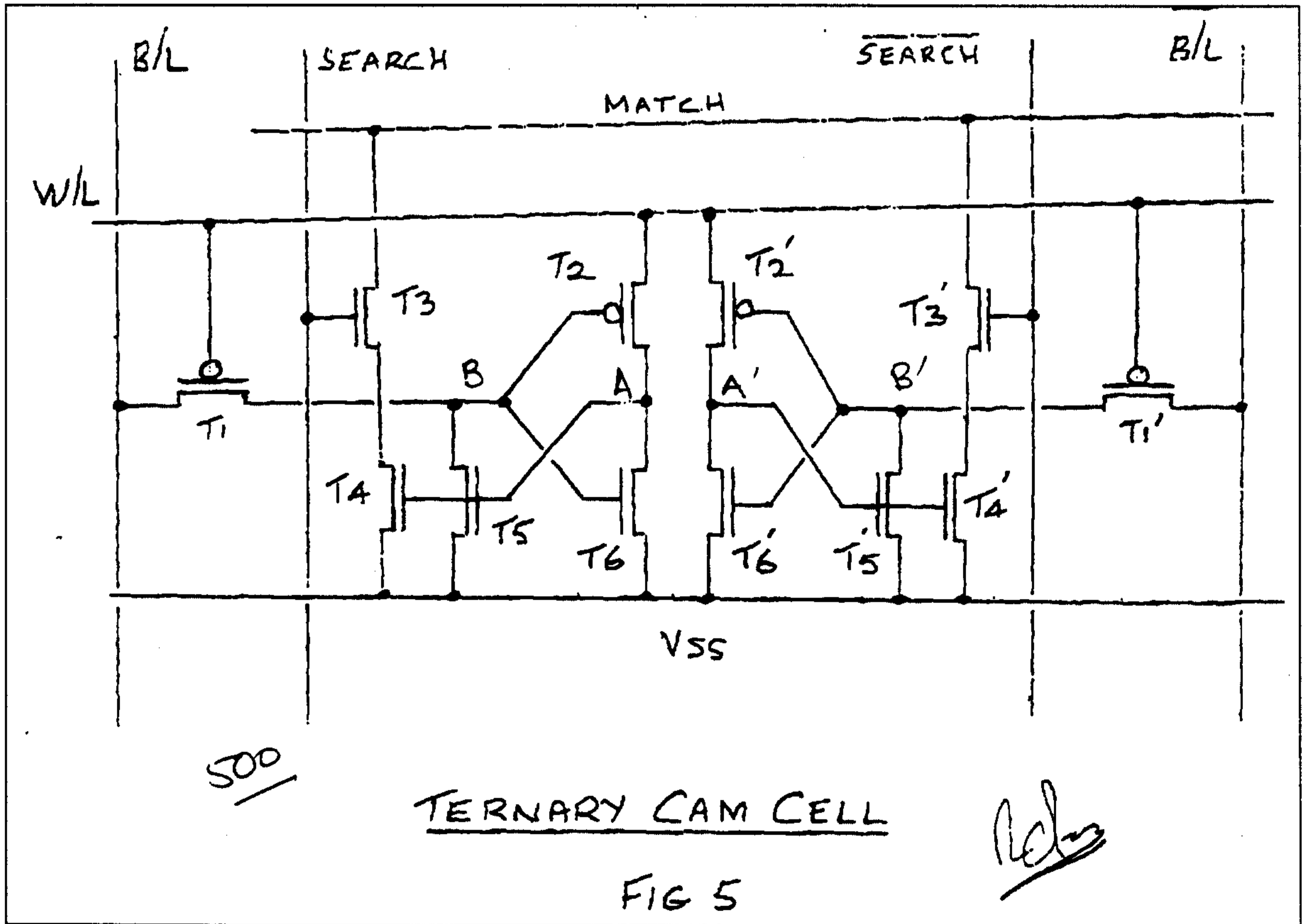
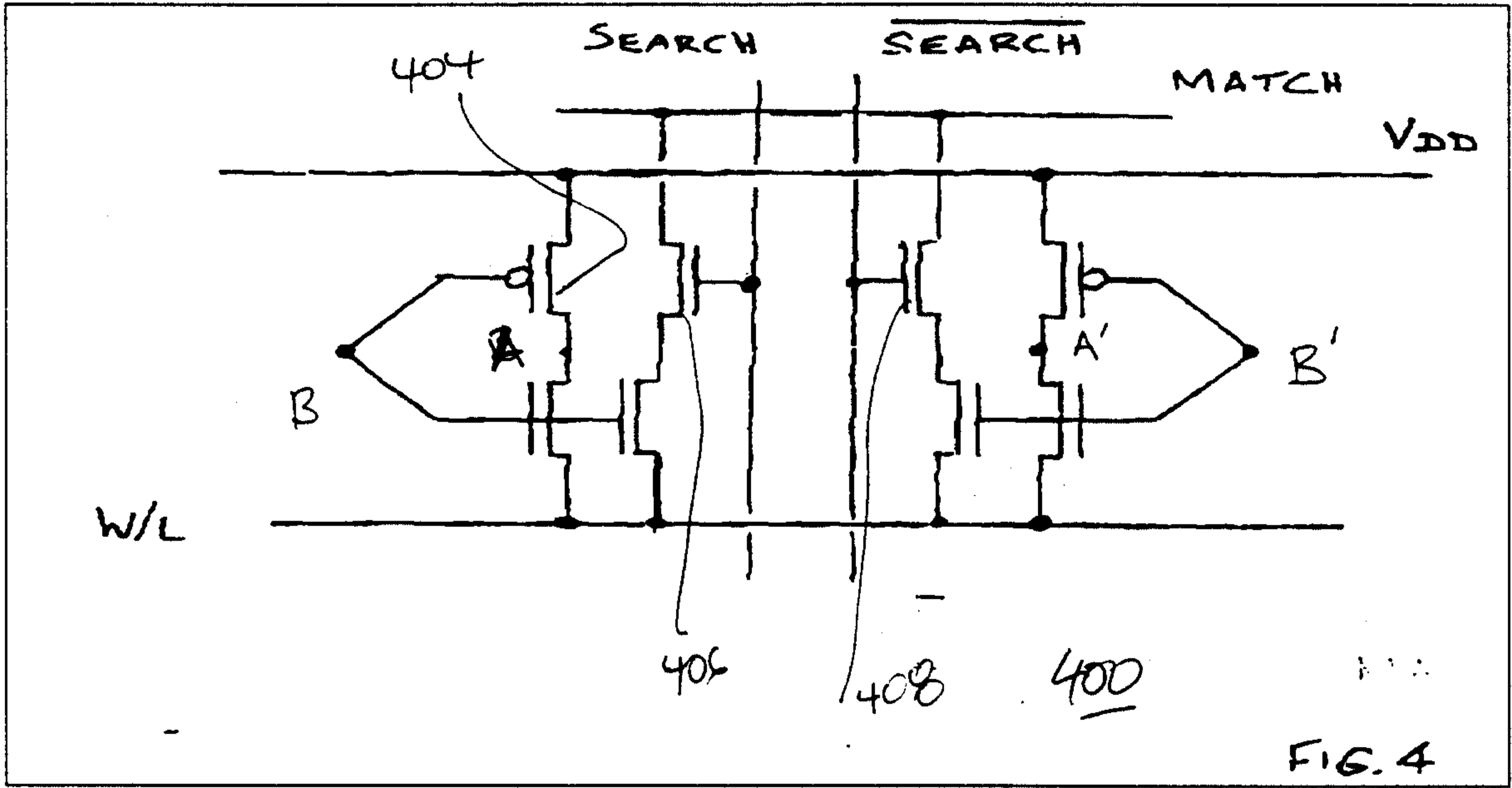


Fig. 1 PRIOR ART.





B/L

200

