A Design for High-Speed Low-Power CMOS Fully Parallel Content-Addressable Memory Macros

Hisatada Miyatake, *Member, IEEE*, Masahiro Tanaka, and Yotaro Mori

***Abstract*—Described is a design for high-speed low-power-consumption fully parallel content-addressable memory (CAM) macros for CMOS ASIC applications. The design supports configurations ranging from 64 words by 8 bits to 2048 words by 64 bits and achieves around 7.5-ns search access times in CAM macros on a 0.35-μm 3.3-V standard CMOS ASIC technology. A new CAM cell with a pMOS match-line driver reduces search rush current and power consumption, allowing a NOR-type match-line structure suitable for high-speed search operations. It is also shown that the CAM cell has other advantages that lead to a simple high-speed current-saving architecture. A small signal on the match line is detected by a single-ended sense amplifier which has both high-speed and low-power characteristics and a latch function. The same type of sense amplifier is used for a fast read operation, realizing 5-ns access time under typical conditions. For further current savings in search operations, the precharging of the match line is controlled based on the valid bit status. Also, a dual bit switch with optimized size and control reduces the current. CAM macros of 256×54 configuration on test chips showed 7.3-ns search access time with a power-performance metric of 131 fJ/bit/search under typical conditions.**

***Index Terms*—ASIC, associative memory, CAM, CAM cell, CAM macro, content-addressable memory, low power, match line, memory macro, sense amplifier.**

I. Introduction

CONTENT-ADDRESSABLE memory (CAM), especially fully parallel CAM, provides a unique exclusive fast data-search function by accessing data by its content rather than its memory location indicated by an address. This CAM function can support a wide range of applications such as lookup tables, databases, associative computing, data compression, and others [1]–[9]. Recently in the network computing era, fast lookup tables are required for network address resolution in network switches and routers such as LAN bridges/switches [10], ATM switches, and layer-3 switches, and CAM's fast search functions are especially useful in supporting the quality of service (QoS) required for real-time applications like multimedia data transmission. Even faster search operations are desired for higher speed communications networks like OC-192 and OC-768 where address resolution within less than 10 ns is required.

In spite of the powerful search function, the use of CAMs is limited to applications where high memory capacity is not the deciding factor because of its relatively low memory density compared with simpler memories like DRAMs and SRAMs and of its relatively high cost resulting from that. CAM's low memory density is mainly due to its area-consuming memory cells and the difficulty of implementing the column address [11]. Another problem to be considered is that the large surge current and large power consumption in the search operation, which come from the inherent nature of parallel CAMs' parallel search, make it difficult to densely integrate a large number of memory cells. For the fully parallel search operation, the input reference data which is being searched for must be sent to all memory locations, and all cell arrays are activated for data comparisons at the same time. During the search operation, therefore, current flows in major data system circuits including long heavy data lines and all bitlines in all cell arrays that are also heavy because a large number of cells are connected. Current also flows in all match detection circuits that include heavy word match lines and are implemented for each data word, and in other search-related circuits such as encoders and hit/miss signal generation circuits. One of the techniques to reduce the search current in the match circuits is to use a NAND-type match-line circuit to get a word match signal [12], rather than the NOR-type circuit generally used for speed-oriented CAMs. In the NAND-type match-line technique, the match-line driver devices in the cells belonging to a word are connected in series. Not only is the NAND-type circuit inherently slower than the NOR-type circuit, but also the former can be much more strongly influenced in the circuit speed than the latter by wafer process variations.

This paper describes a design applied to a series of fully parallel CAM macros for standard CMOS ASIC chips without any special devices such as low-V_{th} devices. For high-speed search, the design uses a wired-AND (NOR-type when used with nMOSFETs) match-line structure with pMOS match-line drivers in the memory cells and a match sense amplifier. They also, together with some other techniques, reduce the search current for less noisy circuit operations and lower power consumption to support fast search. Typical macros constructed using 0.35-μm 3.3-V technology based on this design show 6.3- to 8.8-ns search access times including the address encoding for the matched memory location with priority given to the smallest address. Some of the circuit techniques for the search operation in this design are also effective for increasing the speed and reducing the current in read and write operations, while some other architectural and circuit considerations are provided specifically for these operations, too. Section II describes the logical and physical architectures commonly used for various configurations of the macros. Section III introduces the memory cell and the CAM word comprising the cell, and discusses their advantages. Section IV shows the

Manuscript received August 22, 2000; revised January 10, 2001.

The authors are with the Yasu Technology Application Laboratory, IBM Japan Ltd., Shiga-ken 520-2392, Japan (e-mail: hmiyatak@jp.ibm.com).

Publisher Item Identifier S 0018-9200(01)04128-2.

0018–9200/01$10.00 ©2001 IEEE

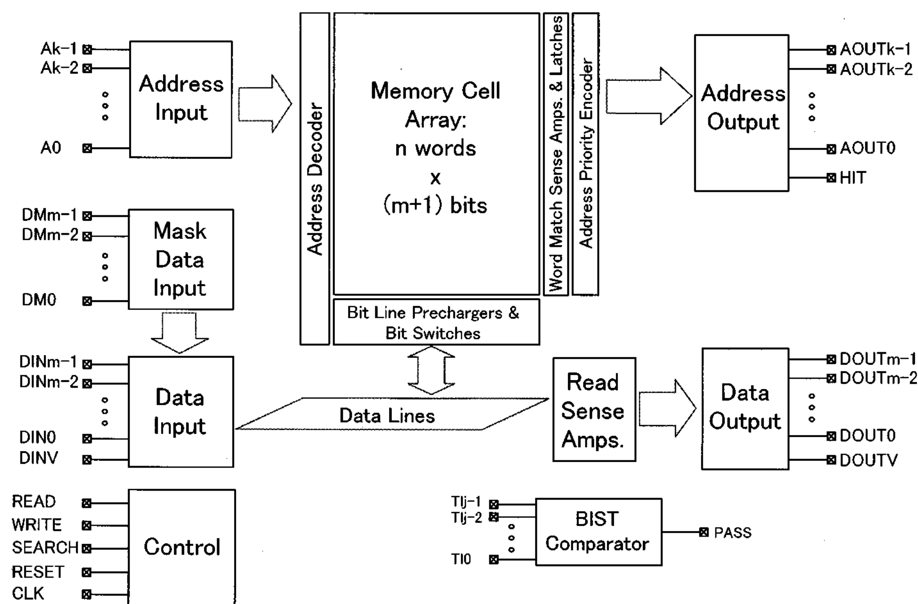


Fig. 1. Logical functional block diagram of the CAM.

sense amplifiers with a latch function used for high-speed word match signal sensing and read data sensing with minimized current. Section V explains further power saving techniques applied to match-line precharging. Section VI presents a bit switch structure and its adaptive control best fitted for read, write, and search operations, and Section VII briefly discusses the power savings in search operations. Section VIII gives a brief description of the technology in which the design is implemented, some circuit simulation and hardware data regarding performance, and physical design data including sizes.

II. Logical and Physical Architectures

The design approach was taken so as to allow the use of as many common circuits and circuit blocks as possible in all supported configurations.

The logical functional block diagram of the developed CAM macros is shown in Fig. 1. In search operations, the reference data are driven to and compared at all locations in the memory cell array at the same time. The memory cell array of $n \times m$ CAM macro logically consists of n words by $(m+1)$ bits including a valid bit, where n may vary from 64 to 2048 with a step of 64 and m from 8 to 64, both depending on the actual implementation of the design. For each data bit, a mask bit is provided to be used to mask or suppress the search and the write operations for that bit. The mask data are acquired in the same cycle as the search or write cycle, and thus no extra cycles are required for setting the mask data. Each word has the valid bit which indicates whether the data stored in the word is to be used in search operations. A match line, a match sense amplifier, and a match latch are associated with each word. The match/nonmatch information in the match latches is fed to a priority encoder to return the smallest address of all matched words. A hit/miss signal is also generated, its "high" state indicating a "hit" or "multiple hits" and the "low" state indicating "no hits" or "miss". A reset command is provided to reset all valid bits to the invalid state in one cycle. The data-in port, including the valid bit input port, is used for write operations as well. The ordinary write and read operations like those in standard RAMs are executed by supplying a write and a read command respectively to the corresponding input port and an address to the address input port. An operation, which is a search, write, read, or reset, is triggered by a clock signal's falling edge at which time all input signals are sampled. The search results and the read-out data come out at their output ports after their respective access times.

Fig. 2 shows the floor plan or physical architecture of a common building block [Fig. 2(a)] and typical configurations [Fig. 2(b) for $64 \times m$ CAM, (c) for $256 \times m$ CAM, and (d) for $2\text{K} \times m$ CAM] of the CAM macros. The memory cell array is divided into subarrays, each of which comprises 64 words. The match signal sense amplifiers and match latches, and a 64-to-1 priority encoder are placed at the side of the subarray. The 64-to-1 priority encoder provides a hit/miss signal and the smallest address in binary code of the matched words in the 64-word subarray. The local wordline drivers are placed at the other side of the subarray. The bitline precharge devices and the bit switches reside below the subarray. These circuit blocks and their local control circuits form a common building block for various configurations with the same word length. To change the word length, the necessary number of 64×1 cell blocks with bit switches and bitline precharge devices are added to or removed from the subarray. In CAM macros containing more than 64 words, the outputs from the 64-to-1 encoder are further encoded among the neighboring common building blocks and sent to a central encoder. The bitlines in the subarrays are commonly used for read, write, and search operations, and are connected to the global bidirectional data lines through the bit switches. The read sense amplifiers and the data drivers are centralized, located with the global control circuits and input/output circuits. Thus the read sense amplifiers are single stage and shared among a maximum of 16 subarrays to save

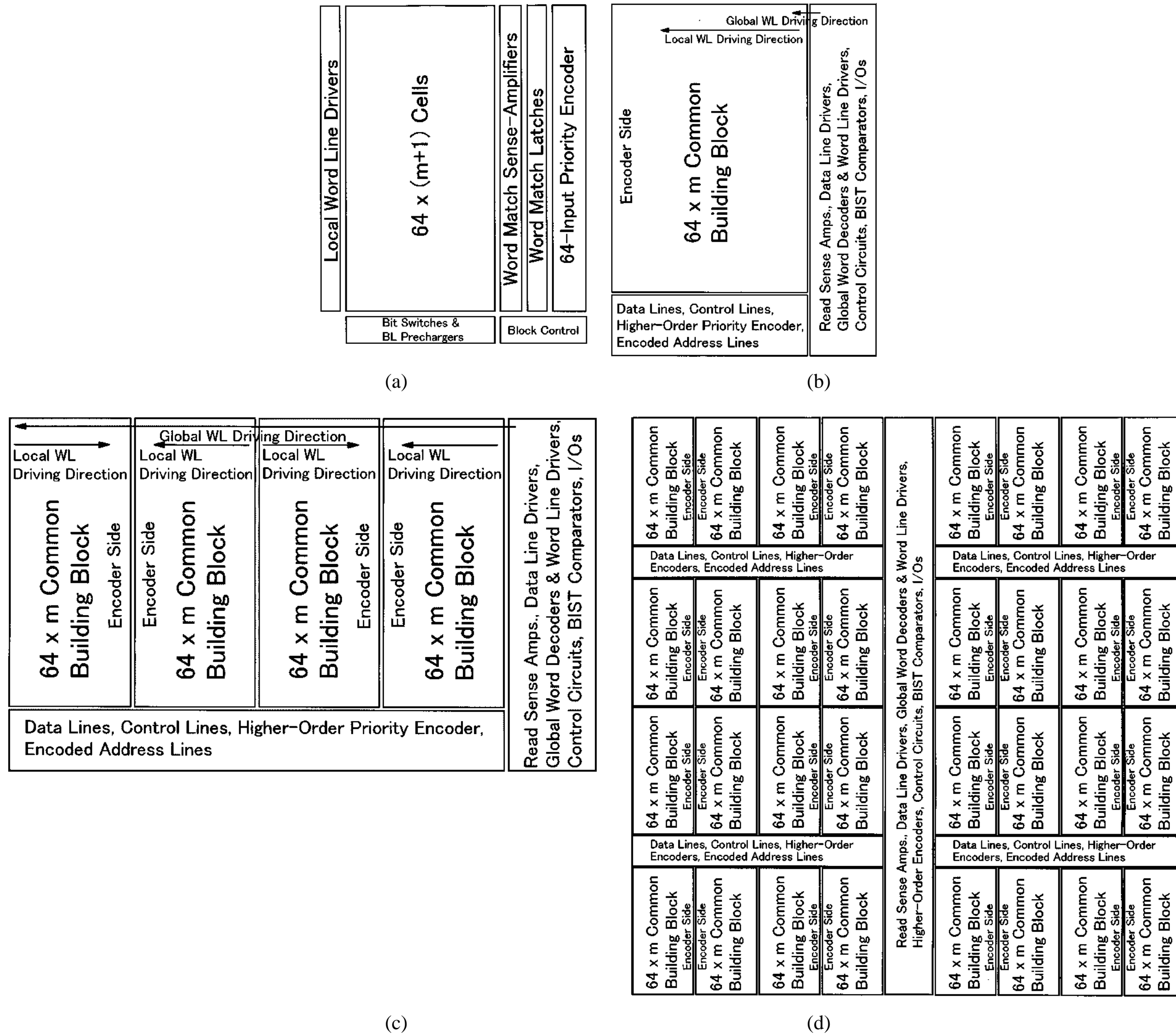


Fig. 2. Floor plans. (a) $64 \times m$ common building block for $n \times m$ CAMs. (b) $64 \times m$ CAM. (c) $256 \times m$ CAM. (d) $2048 \times m$ CAM.

silicon area and electric current. In write and search cycles, the data line drivers drive the data lines to a maximum of eight subarrays according to the input data and the mask data. Global wordlines run over a maximum of four subarrays. A set of global wordline decoders and drivers support a maximum of four subarrays and only one set of them is activated at one time according to the input address. The local wordline drivers receive select signals from the global wordlines and a subarray select circuit to drive a selected local wordline in a selected subarray for read and write operations.

The number of words per subarray, which is 64, was kept rather conservative because a nine-transistor cell and the common bitline scheme were adopted. This will be discussed in the latter half of Section IV-A concerning the sense amplifiers.

III. Memory Cell and Word Structure

The CAM cell developed for this work is a nine-transistor cell as shown in Fig. 3(a). The cell incorporates an ordinary six-transistor SRAM cell to store a bit of datum, a comparison circuit containing two nMOSFETs, and a pMOSFET word match-line driver device which may be called a bit-match device. A major difference from a conventional nine-transistor CAM cell with active pull-down [13], shown in Fig. 3(b), is that the pull-down device to drive the match line to a low level is a pMOSFET rather than an nMOSFET. The comparator devices' gate nodes are connected to the storage nodes in such a way that the match-line driver pMOSFET is off when a match is detected in the cell, which is the same with the conventional nMOS pull-down device cell.

The CAM word is structured as shown in Fig. 4, comprising the cell presented above. The bit-match devices from the memory cells in the word are connected in parallel to the word match line to form a dynamic wired-AND, i.e., a NOR-type, match-line structure for fast search access, similar to a conventional CAM word using an nMOS pull-down driver cell. The word match line stays at its precharge state when a word match is detected. With the intention to use a sense amplifier in order to detect the small signal on the match line, the precharge level was selected so as to be a level suitable for high-gain operation of the sense amplifier presented in the next section. As the sense amplifier has an nMOS gate input, the desirable precharge level is a high level rather than a low level. However,

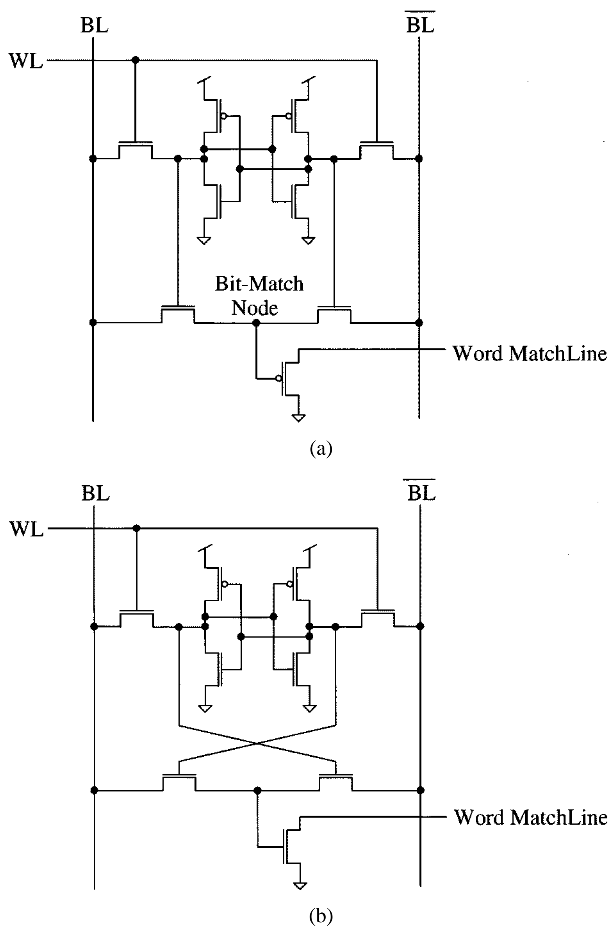


Fig. 3. CAM cells. (a) This work. (b) Typical nine-transistor nMOS pull-down device cell.

to reduce the voltage swing at the same time, the match line is precharged through an isolation nMOS transistor. Thus the precharge level is $V_{\rm thn}$ below $V_{\rm DD}$, where $V_{\rm thn}$ is the threshold voltage of the deeply back-gate-biased nMOSFET and $V_{\rm DD}$ is the power supply voltage. The match sense node is precharged to the full $V_{\rm DD}$ to allow the use of the power supply voltage as the reference voltage for the sense amplifier. As the match line is precharged to the high state, which is the same as the match state, the search access time is determined by the detection of the word nonmatch state, the lower state. The stray capacitance of the sense node is much smaller than that of the match line, and therefore the sense node discharging is accelerated before being amplified by the sense amplifier [14]. Using a high-speed sense amplifier with the isolation nMOSFET, instead of a primitive logic gate, for sensing the match-line signal makes it possible to attain a fast search operation over the wide range of word lengths even with a reduced voltage swing, thus a smaller current. In a typical conventional CAM word structure using an nMOS pull-down device like the cell in Fig. 3(b) and with a primitive gate driven by the match line, the word match state is given by the power-supply voltage level. Thus usually the match line is precharged to $V_{\rm DD}$ before or at the start of a search operation and discharged to the ground level for a nonmatch [4], [5], [13], [15]. The pMOSFET match-line pull-down driver is weak in driving down the word match line in nonmatch case compared with nMOSFET. The wired-AND match-line structure with the nMOS isolation device and sense amplifier was adopted to minimize the effect of this pMOSFET disadvantage. The activation timing of the sense amplifier was determined with the case where only one bit in the word has nonmatch state.

Advantages brought by having a pMOSFET instead of an nMOSFET as the match-line pull-down device are as follows.

1) The match-line discharge level is a level elevated by the threshold voltage of the deeply back-gate-biased pMOSFET, not a full low level or the ground level, leading to smaller rush current, smaller capacitive coupling noise, and power savings through voltage swing reduction.
2) The bit-line precharge level is compatible between the read/write and search operations, specifically a high level for all operations.
3) The match-line precharge or standby level is also compatible between the read/write and search operations, a high level for all operations.
4) The bit mask signal level on the bitlines is compatible between the search and write operations, a high level on both true and complement bitlines in both cases.
5) As a result of 2) and 4), the bit mask signal level on the bitlines is always the same as the precharge level regardless of the requested operations.

Advantage 1) reduces the match-line voltage swing toward the lower state by about 0.7 to 1 V. Along with the reduced precharge level of the match line as previously described, the resultant voltage swing is $V_{\rm DD} - V_{\rm thn} - |V_{\rm thp}|$, where $V_{\rm thp}$ is the threshold voltage of the deeply back-gate-biased pMOSFET. If $V_{\rm thn}$ and $|V_{\rm thp}|$ are both denoted by $V_{\rm th}$ for the sake of simplicity because they are close to each other, the match-line voltage swing is $V_{\rm DD} - 2V_{\rm th}$. Since all nonmatching match lines are discharged in a search operation in case of the NOR-type match-line architecture, and since most words store nonmatch data in typical cases, this reduction of the discharge current is nearly proportional to the word count or the address depth. The discharged match lines must be charged up during the precharge period. Thus the total match-line charging and discharging current is reduced roughly by a factor of $2V_{\rm th}/V_{\rm DD}$ when compared with the case of the rail-to-rail ($V_{\rm DD}$ to the ground level) voltage swing. When $V_{\rm DD}$ is 3.3 V and $V_{\rm th}$ is 0.7 V as a modest value for a deeply back-gate-biased $V_{\rm th}$, this factor is 0.42, i.e., more than 40% of the match-line current is saved. The noise level is reduced by the same factor to the first order approximation, contributing to more stable circuit operation. The match-line charging and discharging power consumption is also proportional to the voltage swing, therefore more than 40% of the match-line power consumption is saved ($P_{\rm d,charge/discharge} \propto f \cdot C_{\rm L} \cdot V_{\rm S} \cdot V_{\rm DD}$, where f is the frequency of the operation, $C_{\rm L}$ is the capacitance of the node to be charged and discharged, and $V_{\rm S}$ is the voltage swing of the node, and only $V_{\rm S}$ is reduced with $V_{\rm DD}$ not changed). The match-line charging and discharging power consumption occupies about 23% of a macro total power consumption when implemented in a 0.35-μm technology described in a later section, therefore about 9.7% of the total power consumption is saved in this technology.

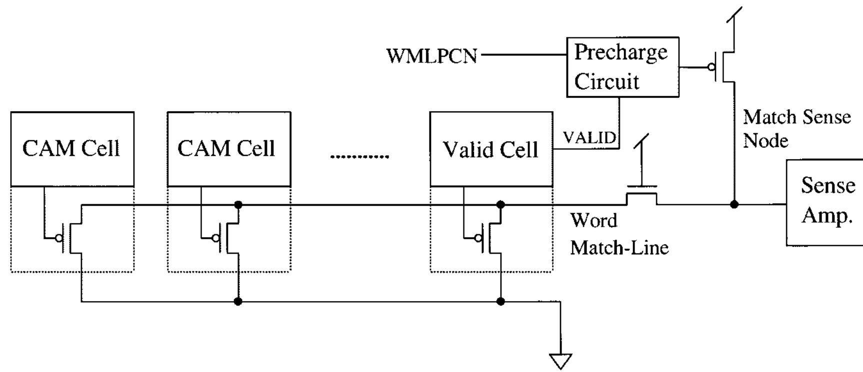


Fig. 4. CAM word circuit. Bitlines, wordlines, and sense amplifier's common node are not shown. WMLPCN is a global precharge signal (active "L") for word match lines.

TABLE I
PRECHARGE LEVEL AND MASK LEVEL CONSISTENCY AT BITLINES AND WORD MATCH LINES

CAM Cell	Precharge for WRITE/READ			Precharge for SEARCH			Write Mask per Bit		Search Mask per Bit	
	Bit Line	$\overline{\text{Bit}}$ Line	Match Line	Bit Line	$\overline{\text{Bit}}$ Line	Match Line	Bit Line	$\overline{\text{Bit}}$ Line	Bit Line	$\overline{\text{Bit}}$ Line
This Work	H	H	H	H	H	H	H	H	H	H
Conventional	H	H	L	L	L	H	H	H	L	L

Advantages 2) and 3) enable the CAM to be ready in a single standby state for both read/write and search operations, which leads to higher speed operations by eliminating precharge operations at the start of the cycles and to lower current operation because there is no standby state change between read/write and search. With the conventional nine-transistor nMOS match-line driver cell in Fig. 3(b), the bitlines must be discharged and then the match line must be precharged when a search operation is requested if the default standby state is for a read or a write operation [13]. Conversely, if the CAM is ready for a search operation in its default standby state, then the bitlines must be precharged to a high level and the match line is thereby discharged when a read or write operation is requested. Note that both bitlines and match lines impose heavy capacitive load on their drivers and prechargers, thus on power lines.

Advantage 4) simplifies the bit mask circuits and reduces current because no mask signal polarity change is required between search and write. With the conventional nMOS pull-down device, the bitlines and also the data lines must be at a low level for bit masking in a search cycle [4], [13], [15], and at a high level for bit masking in a write cycle.

Advantage 5) reduces the current and contributes to higher speed as the bitlines and data lines need not change their states from the precharge level to the mask level when a bit mask is requested. With the conventional cell, the bitlines and data lines must discharge to a low level for search bit masking, if the default standby state is for a read or a write operation.

Table I summarizes the comparison of the states between the new cell and the conventional cell discussed above.

The nine-transistor static CAM cells with bitlines common to write, read, and search operations, like one proposed in this paper, are superior in their space and wiring layer requirements to other types of static cells [13]. Especially in ASIC applications, it is a desired practice to avoid the use of higher level wiring layers in macros in order to ease the chip-level global wiring task.

Note that lowering the match-line precharge level to $V_{\rm DD} - V_{\rm thn}$ is quite compatible with the cell's bit-match node's precharge and high level that is also $V_{\rm DD} - V_{\rm thn}$. If the word match line is precharged up to $V_{\rm DD}$, the bit match pMOSFET is at its marginal state of ON and OFF, and therefore the source–drain leakage current flows through the pMOSFET in every cell in the precharge and match states, leading to larger standby current.

IV. SENSE AMPLIFIERS

The sense amplifiers designed for read data sensing and for match-line signal sensing in the CAMs are shown in Fig. 5(a) and (b), respectively. These amplifiers have the same circuit topology, but one is used as a differential amplifier with symmetric device sizes [Fig. 5(a)] and the other is used as a single-ended amplifier with asymmetric device sizes at the input stage and with a reference voltage applied to one of the input nodes [Fig. 5(b)]. The symmetric one has its predecessor aimed at SRAM application [16]. The sense amplifiers used in the CAMs have two stage drivers for the SET node shown in Fig. 5 to avoid potential signal collapse at the early stage of the amplification, and output buffers to speed up the output signal. The asymmetric amplifier has an equalizer device between its internal true and complement nodes to get more stable sense threshold voltage margin. The main part of the amplifiers consists of two primitive amplifier circuits; one is the well-known cross-coupled inverters and has a positive feedback amplification function and

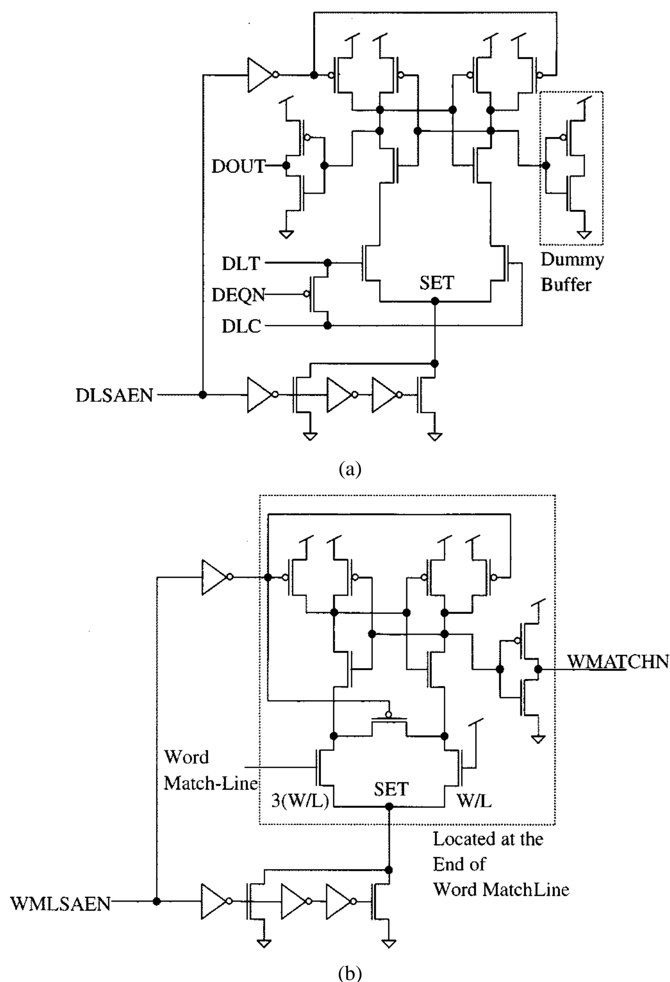


Fig. 5. Sense amplifiers for (a) read data sensing and (b) match-line signal sensing.

a latching function, while the other is used as an input-stage preamplifier. This circuit structure leads to the following useful characteristics common to the two amplifiers:

1) high sensitivity and high speed amplification because of the separation of inputs and outputs and the lightly loaded, buffered outputs, in addition to the inherent high performance resulting from the circuit structure;
2) rail-to-rail output voltage levels;
3) no direct current flow from the power supply line to the ground line except during the state transition period which is very short;
4) a latching function which holds the amplified signal even after the input signals have disappeared as long as the input nodes are at the voltage levels at which the input-stage devices are ON and the sense amplifier is kept enabled;
5) the sensing speed which has little dependency on the input signals' absolute voltage levels, as shown in Fig. 6 (top and center), which illustrates the case of differential read sense amplifier.

A. Read Data Sense Amplifier with Latch Function

The differential amplifier can quickly amplify a differential read signal as small as 10 mV of effective voltage, as shown in the circuit simulation waveforms in the bottom graph of Fig. 6. This amplifier has a dummy output buffer to preserve the symmetry. Characteristics 2), 3), and 4) above make this amplifier superior to other active load differential amplifiers such as the current-mirror load amplifier widely used in SRAMs [17], [18]. The high sensitivity, high speed, and rail-to-rail amplification capability enable the CAM macros to use a single-stage centralized sensing scheme, as shown in Figs. 1 and 2, with small signal swings on the global data lines as well as on the bitlines. Characteristic 4) eases the timing design of the data line and bitline precharge signals, i.e., enables the early starting of the precharge after amplification, and thus contributes to shortening the cycle time. Characteristic 5) makes it possible to select the bitline and data line precharge levels from a wide range of voltage levels, e.g., $V_{\rm DD} - V_{\rm thn}$ for both the bitlines and the data lines as used in this design, and thus to reduce their charging and discharging current, leading to lower noise level and smaller power consumption.

One thing to pay attention to in using a nine-transistor CAM cell along with the common bitline scheme is that the bitline load varies depending on the data stored in the cells connected to the bitline pair. This load difference affects the read operation most severely. For example, assume that the cell design in Fig. 3(a) is used, that one cell stores "0" which is represented by an "L" level on the true side and an "H" level on the complement side, and that all the other cells connected to the same bitline pair have "1." Under this situation, when the datum is being read out from the cell storing "0," that cell's driver transistor must drive the true bitline to which the bit-match devices of all the other cells are connected through their compare devices. In this case, the true bitline is heaviest. The same situation occurs even with the cell of Fig. 3(b) with another data combination. The high-sensitivity high-speed sense amplifiers, the small word count of the subarrays of memory cells which was noted at the end of Section II, and the high-conductance bit switch described in Section VI make it easy to cope with this situation. Relatively short and thus light bitlines connected to the data lines through the bit switches reduce the effect of the load difference above. Worst-case simulations (3.3 V/125 °C/worst process) run on the circuit implementation into a 256×54 CAM macro on the technology described in Section VIII shows that the heaviest case differential signal on the data line pair is 225.21 mV at the sense timing and that the lightest case 238.53 mV. The heaviest case differential signal on the bitline pair is 304.24 mV at the same timing while the lightest case is 326.83 mV. The sense amplifier has a large enough margin for the data line signal that it must amplify.

B. Match-Line Signal Sense Amplifier with Latch Function

The sense amplifier for match-line signal sensing has an asymmetric imbalanced driving capability at the input-stage devices [Fig. 5(b)]. The input gate node of the weaker device is connected to the power-supply voltage as a reference voltage, requiring no reference voltage generator. The difference in the driving capability is implemented by using different channel width over channel length (W/L). The driving capability could be changed also by using different threshold voltages, but this requires a special $V_{\rm th}$ transistor which in turn would require wafer process change. In contrast, using different W/Ls can

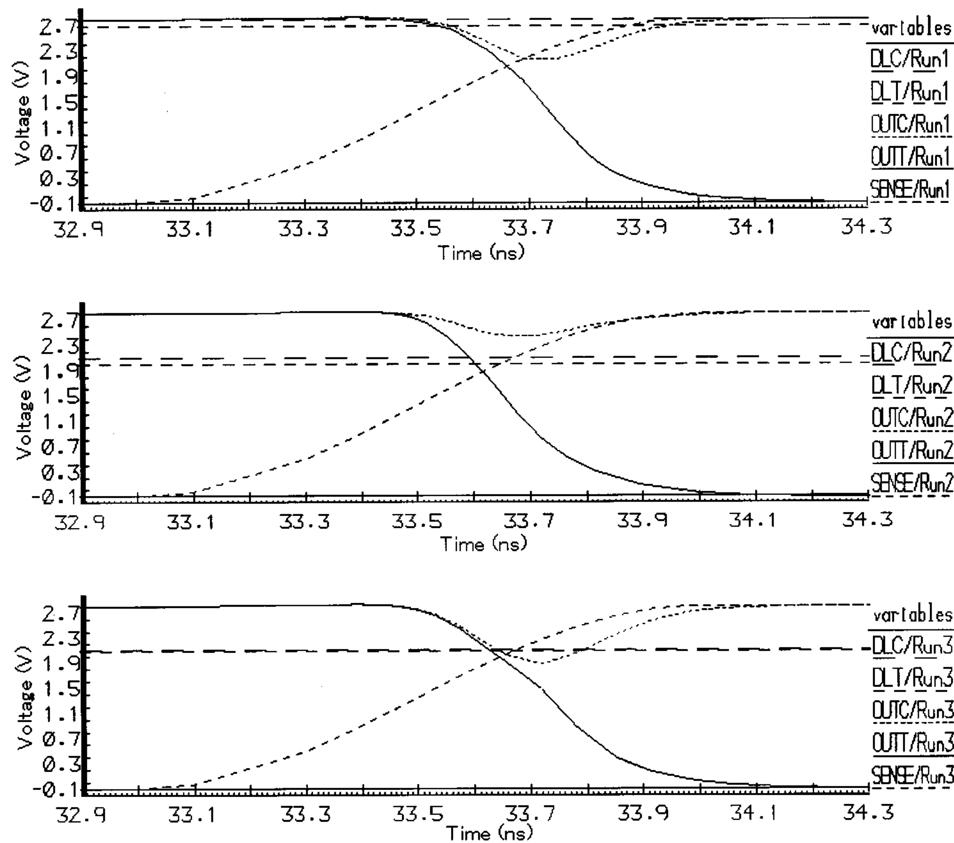


Fig. 6. Sense amplifier's input (DLT and DLC) and output (OUTT and OUTC, both of which are inputs to the output buffers) waveforms with sense-amplifier enable signal (SENSE, which is the input to the SET node driver) for 100-mV (top and center) and 10-mV (bottom) differential inputs at 2.9-V $V_{\rm DD}$ and 125 °C.

be done only within circuit design. The resultant sense-latch amplifier is a single-ended amplifier suitable for match-line signal sensing. The ratio of W/L between the input side and the reference side transistors was determined through circuit simulations for the target technology described in Section VIII. The optimum ratio was searched for varying the channel widths and lengths of the input-stage nMOSFETs and other FETs, the power supply voltage, temperature, and wafer process conditions. The influence of the skews, or independent variations, of $V_{\rm th}$s, Ws, and Ls of the devices were also investigated. The effect of having the output buffers was checked and the output buffer was attached only to the reference side. The match-line load, the leakage current through the pMOSFET match line drivers in the cells, and the isolation nMOSFET size were also varied in total sensing simulations. The sense amplifier's threshold voltage has enough margin over the range of $V_{\rm DD}$ from 2.6 to 6 V. The hardware is functional at $V_{\rm DD}$s at least from 2.4 to 4.5 V, as shown in Section VIII.

By using the high-speed high-sensitivity sense amplifier for the match line instead of primitive logic gates or one of the variations on primitive logic gates [1], [12], it is possible to have fast search access even with long match lines, hence long words such as 64-bit word, where the match line may change its state slowly due to its heavy stray capacitance. This also implies that the sense amplifier may minimize the differences in search access time over a wide range of word lengths.

Since one match-line sensing circuit is required per match line, namely, per word, in fully parallel CAMs, the sense amplifier must be small and consume little current. As shown in Fig. 5(b), the sense amplifier pitch circuit only requires small six pMOS and five nMOSFETs including its precharger and output buffer, and therefore the circuit occupies only a small silicon area. Single-stage amplification to rail-to-rail outputs and no direct current in the steady states brought by this amplifier minimize the area and the total current consumed in match sensing.

As in the case of the read data amplifier, characteristic 4) serves for easing precharge timing design and shortening the cycle time.

V. Match-Line Precharge Control

The data words which do not contain data valid for searching need not be searched. Also, since invalid words must generate "nonmatch," their match lines are wastefully discharged during search access and recharged again in the precharge phase. In this design, to avoid this waste the precharging of the word match line is suppressed and the word match line is statically connected to the ground through the valid bit cell giving a "nonmatch" when the valid bit does not have the "true" value. Fig. 7 shows the precharge control circuit for this purpose which minimizes the additional delay of the precharge signal.

Suppressing the match-line precharging for the invalid words cuts the current related to their match lines whose stray capacitances are large for wide data CAMs. The power savings by using this technique depends on how many invalid words exist in the CAM, hence on the CAM's usage, and therefore varies

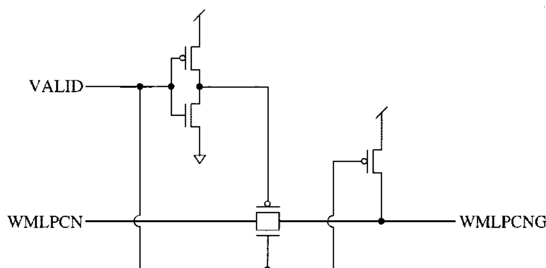
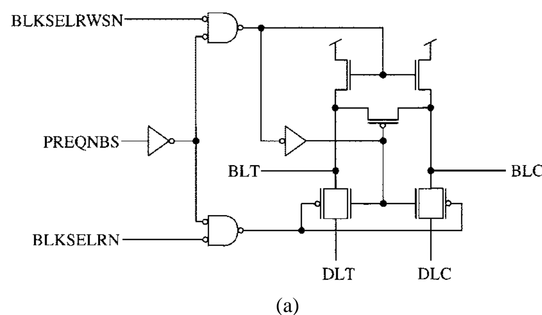


Fig. 7. Match-line precharge control circuit. VALID is from the valid cell, WMLPCN is a global precharge signal (active "L") for word match lines, and WMLPCNG is a local precharge signal (active "L") for a word match line.

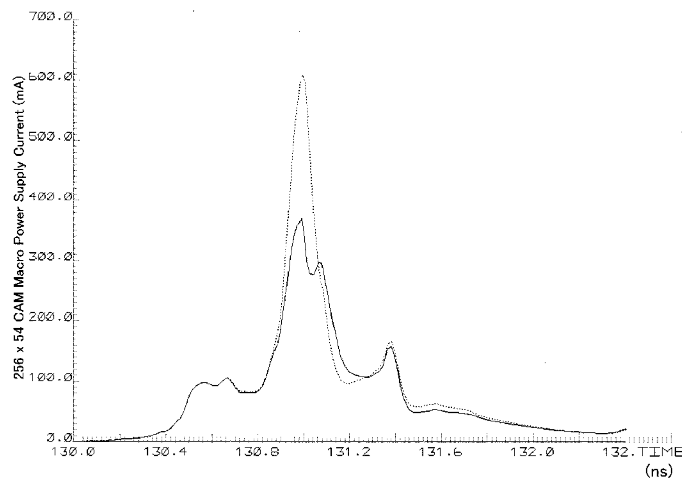
over time. In cases where all valid bits are reset first at the beginning of a task and then valid data are written into the CAMs it is expected that many words remain invalid for certain period of time. In such cases the average power consumption is significantly lower with this scheme of precharge suppression, leading to lower chip power dissipation and lower junction temperature.

VI. Dual Bit Switch and its Control

An nMOS bit switch has enough driving capability, or low enough ON-resistance, for search and write operations in which the low-side data line and bitline are driven down strongly to the ground level. But often, especially in worst-case semiconductor processing conditions, the switch does not have enough driving capability for read operations unless the device width is rather large, because the lower side bitlines do not strongly go low and the gate overdrive is small. This means that the device width must be determined as required for read operations, and that some part of the power for driving these wide bit switches will be wasted in search and write operations. This waste of power can be avoided if a pMOS bit switch is used in parallel with a smaller nMOS bit switch. Furthermore, the power dissipation, i.e., the current, for bit switch driving is minimized if the pMOS bit switch and the nMOS bit switch are controlled so that pMOS and nMOS bit switches are both turned ON for read operations, but only the nMOS bit switch is turned ON for search and write operations, as shown in Fig. 8(a). This bit switch scheme is especially effective when the centralized read sense amplifiers architecture is used with no local read sense amplifiers placed at the bitlines in each subarray and when the read signals on data lines tend to be small. Not only the total size of active bit switch devices in write/search cycles and that in read cycles to get the same amplitude of read signal but also the total size of devices changing their states in the bit switch control circuit in write/search cycles are reduced, as shown in Table II, in spite that the control is more complicated and that the control circuit is a little bigger. The total size of devices changing their states in bit switches and their control circuits are reduced in read cycles as well as in write/search cycles. In CMOS logic circuits the dynamic power is consumed in charging and discharging the gate and other stray capacitances of the devices and wires which change their states and in the form of short-period direct current from the power supply node to the ground node, principally proportional to device widths with device lengths fixed in most devices, at switching that takes place also at the devices changing their states. Table II reflects all devices that change their states



(a)



(b)

Fig. 8. (a) Bit switches, bitline prechargers, and equalizer, and their control circuit. BLKSELRWSN is active ("L") in read, write, and search cycles, BLKSELRN is active ("L") in read cycles, and PREQNBS is active ("H") during bit switch ON period. (b) Macro total power supply current waveforms at bit switch ON timing; solid line for a macro in this work and the dotted line for a macro with nMOSFET only bit switches and their control circuit with the same read signal amount on data lines.

in driving bit switches, and thus the table gives some indication how large the power saving is. The reduction in macro's total current because of the bit switch design change as implemented in the technology described in Section VIII is shown in Fig. 8(b). The 256×56 CAM macro's total peak current, which is not the bit switch related current only, is reduced about 39% at the time when bit switches are turned ON, which leads to lower noise in the access phase which is important for sensing operations. Note that all bit switches in a parallel CAM must be ON in search operations, and therefore minimizing bit switch current has a big effect on reducing search current.

VII. Power Related to Bitlines and Data Lines in CAMs

With the circuit techniques described in previous sections, the search operation is less current-consuming to attain high speed, but the largest current is consumed in driving the bitlines and the data lines because all the low-side bitlines must be driven at once for fully parallel search operation.

Here again, reducing the voltage swing is an effective method for reducing current flow with any cell structure. But elevating the low level of the bitline signals not only greatly affects the speed of discharging the match line with the cells with pMOS match-line driver as used in this work, but also is very problematic for the search operation with conventional cells using nMOS match-line driver. When a match is expected and the

TABLE II
COMPARISON OF TOTAL SIZE OF DEVICES CHANGING THEIR STATES IN BIT SWITCH AND ITS CONTROLLER BETWEEN THIS WORK AND NMOS-ONLY BIT SWITCH CASE IN CASE OF ×54 COMMON BUILDING BLOCK (ARBITRARY UNIT)

	WRITE/SEARCH cycle								
	Bit switch			Controller			Bit switch & controller		
	p-FET	n-FET	Total	p-FET	n-FET	Total	p-FET	n-FET	Total
n only	8.0	55.0	63.0	477.0	143.0	620.0	917.0	3168.0	4085.0
This work	4.5	26.0	30.5	431.0	129.0	560.0	678.5	1559.0	2237.5
Reduction	3.5	29.0	32.5	46.0	14.0	60.0	238.5	1609.0	1847.5
	43.8%	52.7%	51.6%	9.6%	9.8%	9.7%	26.0%	50.8%	45.2%
	READ cycle								
	Bit switch			Controller			Bit switch & controller		
	p-FET	n-FET	Total	p-FET	n-FET	Total	p-FET	n-FET	Total
n only	8.0	55.0	63.0	477.0	143.0	620.0	917.0	3168.0	4085.0
This work	23.5	27.8	51.3	503.0	166.0	669.0	1795.5	1695.0	3490.5
Reduction	-15.5	27.2	11.7	-26.0	-23.0	-49.0	-878.5	1473.0	594.5
	-193.8%	49.5%	18.6%	-5.5%	-16.1%	-7.9%	-95.8%	46.5%	14.6%

TABLE III
SUMMARY OF THE CMOS ASIC TECHNOLOGY IN WHICH THE DESIGN WAS IMPLEMENTED

Design rule	0.35 μm
Substrate	p-epitaxial on p+ substrate
Gate material	Ti-salicided n+ and p+ polysilicon
Gate oxide thickness	8.4 nm (SiO_2 equivalent)
Diffusion	Ti-salicided
Device isolation	Shallow trench isolation (STI)
Wiring layers	1 local interconnect, and up to 5 levels of metal
Inter-layer connection	W(Tungsten) stud
Power supply voltage	3.3 V ± 0.3 V
Operating temperature range	-55°C to 125°C (junction)

nMOS match-line driver must be off, the nMOS driver can leak current if the low-side bitlines have not been driven to a low enough voltage level. Also, more generally, it is a problem for stable write operations with the SRAM-based cells. In this design, therefore, the high level, which is also the precharge level, of the bitline and data line signals are lowered by the deeply back-gate-biased nMOS transistor threshold voltage. This approach with high-performance sense amplifiers demands no sacrifice in search or write speed, and reduces the current related to the bitlines and to the data lines by the factor of $V_{\rm thn}/V_{\rm DD}$ which is about 0.2 to 0.3 in 3.3-V technology. The performance of the read sense amplifier described in Section IV is hardly affected by the data lines' starting voltage which is the precharge voltage.

It also should be noted that the lowered precharge level of the bitlines does not affect the cell's bit-match device gate node voltage which can have some effect on holding the match-line precharge level, because even if the bitline precharge level is $V_{\rm DD}$, the bit-match node voltage is $V_{\rm DD}-V_{\rm thn}$ due to the voltage drop across the nMOS transistor of the comparator circuit in the cell.

VIII. IMPLEMENTATION

The architecture and the circuit designs described above were implemented in a 3.3-V 0.35-μm five-layer metal standard CMOS ASIC technology without any special transistors like low-$V_{\rm th}$ transistors. Table III summarizes the general characteristics of this technology. The CAM design is deployed in various configurations: 1K $\times$ 54, 1K $\times$ 18, 256 $\times$ 54, 256 $\times$ 22, and 64 $\times$ 32. The circuit design and layout design of the local circuits in the common building block as shown in Fig. 2(a) are common to all configurations except some wire lengths. Some circuit elements of global circuits such as the read sense amplifier, data line driver, address decoders, global wordline decoders and driver, and higher level priority encoders as well as the I/O circuit are basically the same in circuit and layout designs except some final driver stages. Many global circuit

TABLE IV
CAM PHYSICAL CHARACTERISTICS IN A 0.35-μm CMOS ASIC TECHNOLOGY

CAM Macro Configuration	Macro Shape	Size in μm (Maximum Length)	Area in mm^2	No. of Pins	Used Metal Layers (Area % for 3rd metal usage)
64w x 32b	L shape	682.56 x 990.72	0.5589	150	2 layers & 6.72% of 3rd metal
256w x 22b	L shape	1103.04 x 967.68	0.9671	124	2 layers & 6.62% of 3rd metal
256w x 54b	Rectangle	1658.88 x 1036.80	1.7199	220	2 layers & 6.11% of 3rd metal
1024w x 18b	Rectangle	1552.32 x 1843.20	2.8612	122	2 layers & 7.68% of 3rd metal
1024w x 54b	Rectangle	2975.04 x 1935.36	5.7578	230	2 layers & 6.79% of 3rd metal

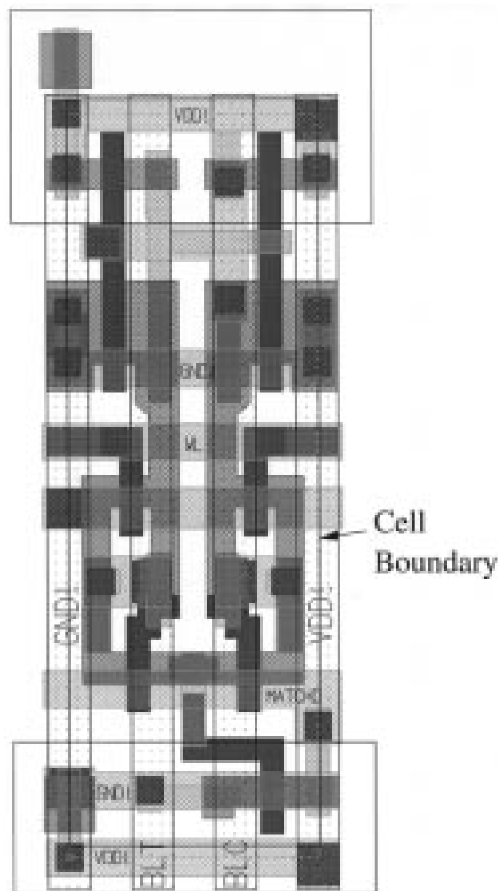
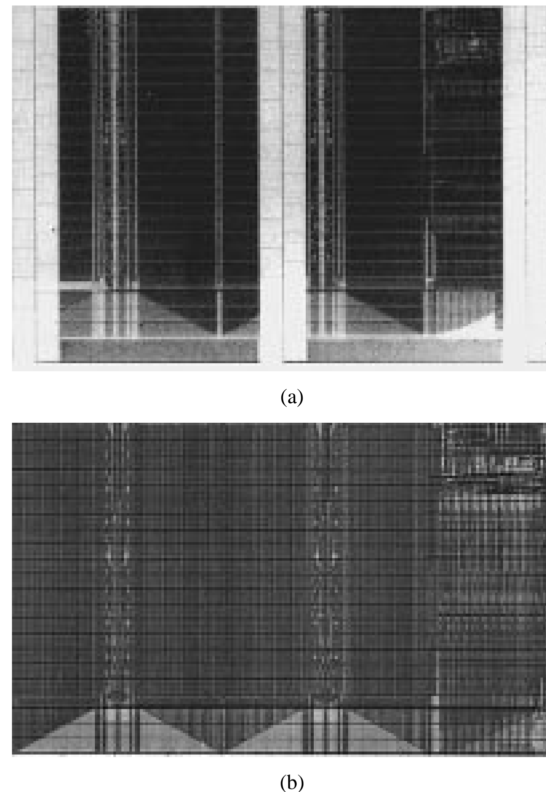


Fig. 9. CAM cell layout in 0.35-μm standard CMOS ASIC technology.



(a)

(b)

Fig. 10. (a) Microphotograph of 256×54 CAM macro in the isolation test chip with power lines added. (b) Plot of the layout design of 256×54 CAM macro.

blocks are also the same in some configurations. The physical sizes of the five configurations are listed in Table IV along with other characteristics of the macros.

The CAMs use two metal layers and a small part of the third metal layer; most of the third metal layer over the CAMs is available for chip global wiring. A kind of buried contact wiring layer is used for local wiring. The nine-transistor cell size is 4.32 μm $\times$ 12.96 μm. No special layout rules are used for the cell physical design. Fig. 9 shows the cell layouts. The bitlines and power lines run vertically on the second-level metal layer, and the global and local wordlines, and the match lines run horizontally on the first-level metal layer, conforming to the ASIC's wiring rules.

Search performances were measured on test chips in which two 256×54 CAM macros were isolated from the other parts of a target chip and pads were added for observation and stimulation. To supply power to the isolated CAMs, wide power buses were added and two pairs of power pads were connected to the power buses. This chip was not intended for the purpose of full characterization, and the macro outputs are not driven by off-chip drivers with large enough transistors. Therefore, precise access times cannot be measured with rather heavy LSI tester loads, but they were obtained by observing the waveform of output signals. Fig. 10 shows a photomicrograph of a 256×54 CAM macro in the test chip and a plot of the layout design of the same CAM macro.

The typical search speed including address encoding with priority, or the search access time, of the 256×54 CAM is 7.3 ns at 3.3 V of the power supply voltage and 40 °C. The access

TABLE V
MEASURED SEARCH ACCESS TIMES OF 256×54 CAM IN NANOSECONDS AT 40 °C (MEAN ± STANDARD DEVIATION)

Case \ V_{DD}	2.4 V	2.7 V	3.0 V	3.3 V	3.6 V	3.9 V	4.2 V	4.5 V
All Non-Match	10.15 ±0.09	8.88 ±0.10	8.06 ±0.10	7.34 ±0.06	6.73 ±0.07	6.42 ±0.05	5.96 ±0.11	5.53 ±0.08
Half Match	10.11 ±0.08	8.92 ±0.09	7.98 ±0.07	7.25 ±0.10	6.73 ±0.09	6.36 ±0.09	5.94 ±0.09	5.53 ±0.10
All Match	9.98 ±0.05	8.80 ±0.07	7.88 ±0.08	7.24 ±0.08	6.66 ±0.07	6.31 ±0.08	5.86 ±0.08	5.50 ±0.06

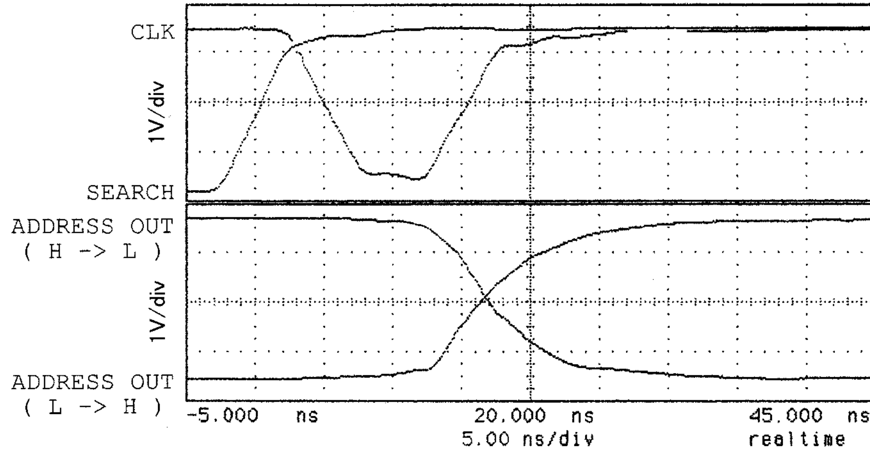


Fig. 11. Waveforms of search command input, clock input, and address output observed on tester probe card.

TABLE VI
ACCESS TIMES AND CYCLE TIMES OF VARIOUS CONFIGURATIONS OF CAMs OBTAINED WITH ASTAP SIMULATION
(3.3 V/25 °C/NOMINAL PROCESS/Output Load = 0.3 pF/Input Transition = 0.25 ns)

CAM config.	Search Access	Search Cycle	Read Access	Read Cycle
64w x 32b	6.25ns	6.80ns	4.60ns	5.44ns
256w x 22b	7.68ns	6.81ns	5.02ns	5.42ns
256w x 54b	7.77ns	6.81ns	5.04ns	5.40ns
1024w x 18b	8.15ns	6.91ns	4.95ns	5.33ns
1024w x 54b	8.80ns	7.17ns	5.07ns	5.56ns

time was measured with nonmatches at all words but one, nonmatches at half of the words, and matches at all words to see the effects of the match lines' discharging current. There were no significant difference among these three cases as shown in Table V, which also shows the measured access times at V_{DD}s from 2.4 to 4.5 V. A typical waveform obtained at tester probe card is shown in Fig. 11. The output transition is slow because the drivers are not large enough for the reason stated above. The distance from the probe points on the card to the chip pads are about 8 cm, and 0.5 ns of delay is calculated for the round trip of this distance. The delays from the 50% point of the input clock edge to the starting point of address output transitions were measured. The output timing point selection is optimistic, but the input timing point selection is pessimistic because the tester-driven input transition time is much larger than that of the on-chip signals. The search access time obtained with ASTAP circuit simulation is shown in Table VI, along with other timing parameters. The read access time and the read cycle time for the same configuration are 5.0 and 5.4 ns, respectively.

The search operating current of the two 256×54 CAMs on the test chips were measured at various search cycle times and power supply voltages in "all word match" case and "all word nonmatch" case. The current of one 256×54 CAM macro are plotted to the search cycle frequency in Fig. 12. The current at 70 MHz of continuous search operations is about 38.5 mA at 3.3-V V_{DD}, or 0.13 W, for the "all word nonmatch" case, which is about 5.5-mA larger than that for the "all word match" case. The difference is about 14%, and this is also the current required to charge up the discharged match lines with the reduced swing voltage. According to the calculations in Section III, the current for match-line swinging should be about 9.6 mA if the match lines swing fully, therefore about 9.6%

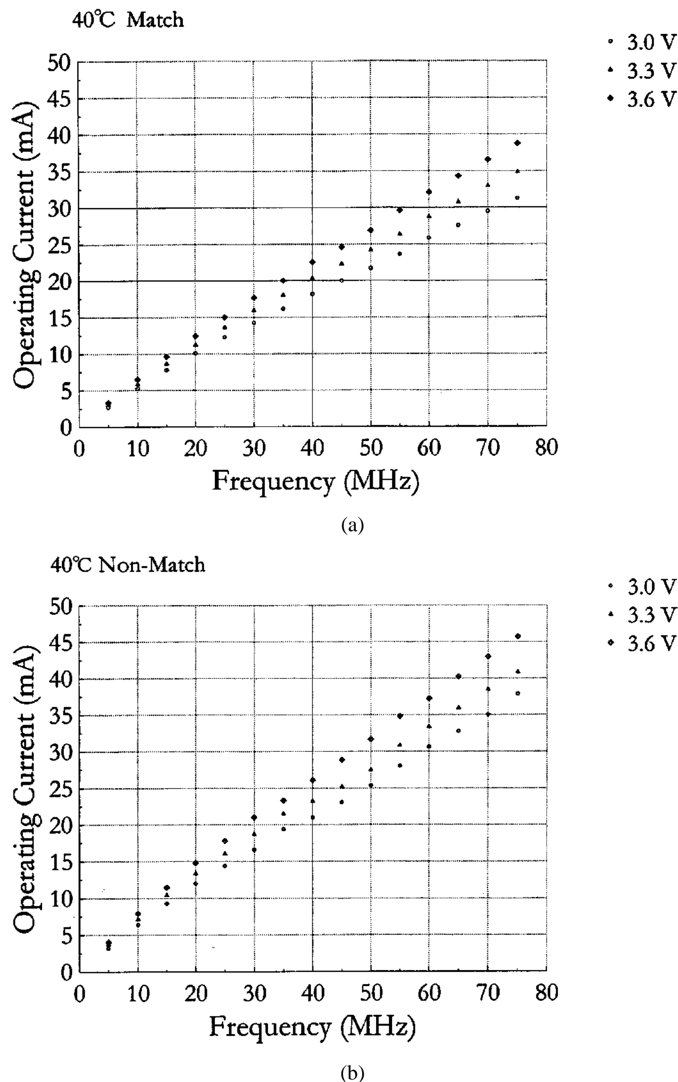


Fig. 12. Measured power supply current versus search frequency for a 256×54 CAM macro. (a) "All word match" case. (b) "All word nonmatch" case.

of the total search power consumption is saved by limiting the match-line voltage swing. The power-performance metric is 131 fJ/bit/search, compared with 270 fJ/bit/search of a previous design using a NOR-type word match-line structure [8] and 83 fJ/bit/search for a NAND-type word match-line structure [19].

Table VI also shows the access times and cycle times obtained by ASTAP simulations of the various configurations of CAMs. It can be seen that their dependence on the word length is relatively small. For example, the access times and cycle times of the 256×54 CAM macro is within 1.2% increase of the 256×22 CAM macro, and the search access time of the 1024×54 CAM macro is about 8% larger than that of the 1024×18 CAM while the former is about 13% larger than that of the 256×54 CAM macro.

IX. Conclusion

A new design for high-speed low-power fully parallel CAM macros has been developed for a standard CMOS ASIC technology. The design supports a wide range of word and bit configurations with small variations of access and cycle times. Using a pMOS match-line driver in the memory cell enables the match-line discharge level to be sustained at the deeply back-gate-biased transistor's threshold voltage above the ground level. The precharge level of the match line lowered by the deeply back-gate-biased isolation nMOSFET further reduces the match-line voltage swing. The cell structure and the reduced current brought about by the reduced match-line voltage swing facilitate the use of a NOR-type (wired-AND) match-line structure suitable for high-speed search even for long words. The cell with the pMOS match-line driver provides other advantages such as bitline precharge levels compatible between the search and the read/write operations, internal data bit mask levels compatible between the search and write operations and with the precharge level, all of which lead to simple control and fast starts of operations and the elimination of the wasteful discharging of the bitlines, the data lines, and the match lines. An unbalanced, rail-to-rail output, single-end latching sense amplifier supports the reduced match-line voltage swing scheme and provides high-speed low-power sensing with the simple reference voltage of $V_{\rm DD}$. A balanced differential sense amplifier of the same circuit topology is used as the data read sense amplifier enabling simple fast centralized single-stage amplification scheme with reduced voltage swing of the bitlines and data lines. The match-line precharging and the match-line logical state are controlled using the valid bit value for the word to further reduce power consumption in search operations. A dual bit switch and its adaptive control were designed for current-saving power-efficient signal transmission between the bitlines and the data lines.

The architecture and design has been implemented in a 0.35-μm CMOS standard ASIC technology, delivering 7.3-ns search access time and 5.0-ns read access time for a 256×54 CAM macro. The search cycle power consumption at 70-MHz successive searches in a 256×54 CAM macro is measured as 0.13 W at 3.3-V $V_{\rm DD}$, giving a power-performance metric of 131 fJ/bit/search.

Acknowledgment

The authors thank R. A. Piro and his team for their helpful suggestions and great efforts in implementing the CAM macros in an ASIC chip and preparing the test chip. They are also grateful to T. Sunaga, T. Hattori, and R. F. Keil for their encouragement and support throughout the development work with difficult targets.

References

[1] T. Takayanagi, K. Sawada, M. Takahashi, Y. Itoh, M. Uchida, Y. Toyoshima, H. Hayashida, and M. Norishima, "2.6 Gbyte/s bandwidth cache/TLB macro for high-performance RISC processor," in *Proc. IEEE 1991 Custom Integrated Circuits Conf.*, 1991, pp. 10.2.1–10.2.4.

[2] K. Nogami, T. Sakurai, K. Sawada, K. Sakaue, Y. Miyazawa, S. Tanaka, Y. Hiruta, K. Katoh, T. Takayanagi, T. Shirotori, Y. Itoh, M. Uchida, and T. Iizuka, "A 9-ns hit-delay 32-kbyte cache macro for high-speed RISC," *IEEE J. Solid-State Circuits*, vol. 25, pp. 100–106, Feb. 1990.

[3] J. P. Wade and C. G. Sodini, "A ternary content-addressable search engine," *IEEE J. Solid-State Circuits*, vol. 24, pp. 1003–1013, Aug. 1989.

[4] K. E. Grosspietsch, "Associative processors and memories: A survey," *IEEE Micro*, pp. 12–19, June 1992.

[5] T. Ogura, J. Yamada, S. Yamada, and M. Tan-no, "A 20-kbit associative memory LSI for artificial intelligence machines," *IEEE J. Solid-State Circuits*, vol. 24, pp. 1014–1020, Aug. 1989.

[6] S. Jones, "Design, selection and implementation of a content-addressable memory for a VLSI CMOS chip architecture," *Proc. IEEE*, vol. 135, pp. 165–172, May 1988.
[7] M. Motomura, J. Toyoura, K. Hirata, H. Ooka, H. Yamada, and T. Enomoto, "A 1.2-million transistor 33-MHz 20-b dictionary search processor (DISP) ULSI with a 160-kb CAM," *IEEE J. Solid-State Circuits*, vol. 25, pp. 1158–1165, Oct. 1990.
[8] T. Ogura, M. Nakanishi, T. Baba, Y. Nakabayashi, and R. Kasai, "A 336-kbit content-addressable memory for highly parallel image processing," in *Proc. IEEE 1996 Custom Integrated Circuits Conf.*, 1996, pp. 13.4.1–13.4.4.
[9] H. Bergh, J. Eneland, and L. Lundström, "A fault-tolerant associative memory with high-speed operation," *IEEE J. Solid-State Circuits*, vol. 25, pp. 912–919, Aug. 1990.
[10] H. Yamada, Y. Murata, T. Maeda, R. Ikeda, K. Motohashi, and K. Takahashi, "Real-time string search engine LSI for 800-Mbit/s LANs," in *IEEE 1988 Custom Integrated Circuits Conf.*, 1988, pp. 21.6.1–21.6.4.
[11] K. J. Schultz and P. G. Gulak, "Architectures for large-capacity CAMs," *Integration, the VLSI Journal*, vol. 18, pp. 151–171, 1995.
[12] K. Schultz, F. Shafai, R. Gibson, A. Bluschke, and D. Somppi, "Fully parallel 25-MHz 2.5-Mb CAM," in *IEEE ISSCC Dig. Tech. Papers*, 1998, pp. 332–333.
[13] K. J. Schultz, "Content-addressable memory core cells: A survey," *Integration, the VLSI Journal*, vol. 23, pp. 171–188, 1997.
[14] L. A. Glasser and D. W. Dobberpuhl, *The Design and Analysis of VLSI Circuits*. Reading, MA: Addison-Wesley, 1985, sec. 5.3.
[15] H. Kadota, J. Miyake, Y. Nishimichi, H. Kudoh, and K. Kagawa, "An 8-kbit content-addressable and reentrant memory," *IEEE J. Solid-State Circuits*, vol. SC-20, pp. 951–956, Oct. 1985.
[16] T. Kobayashi, K. Nogami, T. Shirotori, Y. Fujimoto, and O. Watanabe, "A current-mode latch sense amplifier and a static power saving input buffer for low-power architecture," in *Symp. VLSI Circuits Dig. Tech. Papers*, 1992, pp. 28–29.
[17] K. Ishibashi, K. Komiyaji, H. Toyoshima, M. Minami, N. Ohki, H. Ishida, T. Yamanaka, T. Nagano, and T. Nishida, "A 300-MHz 4-Mb wave-pipeline CMOS SRAM using a multiphase PLL," *IEEE J. Solid-State Circuits*, vol. 30, pp. 1189–1195, Nov. 1995.
[18] B. Prince, *Semiconductor Memories*, 2nd ed. Chichester, U.K: Wiley, 1995, sec. 8.13.5.
[19] F. Shafai, K. J. Schultz, G. F. R. Gibson, A. G. Bluschke, and D. E. Somppi, "Fully parallel 30-MHz 2.5-Mb CAM," *IEEE J. Solid-State Circuits*, vol. 33, pp. 1690–1696, Nov. 1998.

Hisatada Miyatake (M'85) received the B.S. degree in physics from University of Tokyo, Tokyo, Japan, in 1979.

In 1979, he joined the Product Assurance Laboratory (PAL), IBM Japan Ltd., where he was engaged in evaluation and qualification of semiconductor processes, devices, and LSIs and their packaging. In 1980, he worked for vendor LSI qualification in IBM Kingston, NY, and in 1982, for 1-Mb nMOS DRAM evaluation in IBM Burlington, VT. In 1987, he moved to the Yasu site to take part in LSI development activities. His major development work includes design and development of a CRT controller, memory and I/O controller, and plasma display panel controller, all for personal computers such as JX and PS/2, x16 1 Mb nMOS DRAM, DRAM macros for ASICs used for a massively parallel processor chip EXECUBE, 3-D graphics chip, $L2$ cache chips and so on, and CAM macros for lossless data compression applications and for communication applications. Currently, he works on communication LSI and new memory developments.

Mr. Miyatake is a member of the Institute of Electronics, Information and Communication Engineers of Japan, the Japan Society of Applied Physics, and the Physical Society of Japan.

Masahiro Tanaka received the B.S. degree in electrical engineering from Kobe University, Kobe, Japan, in 1990.

He joined IBM Japan, Ltd., at the Yasu Technology Application Laboratory in 1990 and worked on DRAM circuit designs. From 1991 to 1992, he was a member of the low-power DRAM design project at the IBM Burlington Laboratory, Burlington, VT. Until 1996, he was engaged in development of DRAM macro design at Yasu. From 1997 to 1999, he designed a number of content-addressable memories. Since 2000, he has been working on the development of the micro controller.

Yotaro Mori received the B.S. degree in physics from Kyushu University, Fukuoka, Japan, in 1991.

He joined IBM Japan, Ltd., Yasu Technology Application Laboratory, in 1991, and worked on CMOS DRAM circuit design and development. From 1995 to 1996, he was a member of DRAM Macro design group at the IBM Burlington Laboratory, Burlington, VT. From 1996 to 1999, he developed content-addressable memories, their peripheral logic circuits, and built-in test circuits. He is currently working on the development of communication products.