A 60-ns 3.3-V-Only 16-Mbit DRAM with Multipurpose Register

KAZUTAMI ARIMOTO, MEMBER, IEEE, KAZUYASU FUJISHIMA, YOSHIO MATSUDA, MASAKI TSUKUDE, TUKASA OISHI, WATARU WAKAMIYA, SHIN-ICHI SATOH, MICHIHIRO YAMADA, AND TAKAO NAKANO, SENIOR MEMBER, IEEE

Abstract —A single 3.3-V 16-Mbit DRAM with a 135-mm² chip size has been fabricated using a 0.5-μm twin-well process with double-metal wiring. The array architecture, based on the twisted-bit-line (TBL) array, includes suitable dummy and spare word-line configurations which suppress the inter-bit-line noise and bring yield improvement. The multipurpose register (MPR) designed for the hierarchical data bus structure provides a line-mode test (LMT), copy write, and cache access capability. The LMT with on-chip test circuits using the MPR and a comparator creates a random test pattern and reduces the test time to 1/1000. A field shield isolation and a T-shaped stacked capacitor allow the layout of a 4.8-μm² cell size with a storage capacitance of 35 fF. These techniques enable the 3.3-V 16-Mbit DRAM to achieve a 60-ns $\overline{RAS}$ access time and 300-mW power dissipation at 120-ns cycle time.

I. Introduction

POWER SUPPLY voltage is one of the most important design considerations for scaling down ULSI's. There are two possible arrangements. One is an external power supply voltage of 3.3 V, and the other is that of 5.0 V with an on-chip voltage converter. Although almost all 16-Mbit dynamic RAM's [1]–[3] use the second arrangement, the first arrangement [4] has the advantage of lower power dissipation in operation and standby than the 5-V scheme with the on-chip limiter. At the 16-Mbit level, even if the internal voltage limiter is provided to maintain reliability [5], the power dissipation cannot be reduced adequately. Therefore, the 3.3-V power supply is the best choice in meeting an acceptable power dissipation. Also, the 3.3 V enables ideal scaling down to a 0.5-μm level transistor, which brings improved performance while maintaining reliability [6], [7].

This paper describes a single 3.3-V 16-Mbit DRAM which follows the ideal scaling rule. An improved array architecture is required to overcome the small signal and the large coupling noise in a scaled-down DRAM cell array. To solve this problem, an array architecture based on the twisted-bit-line (TBL) [8], [9] technique, and including suitable dummy and spare arrangements, is proposed in Section II. Although a 16-Mbit DRAM has very fast access, testability must be considered from a cost basis. The multipurpose register (MPR) [10] provides high performance and many other useful functions. The main use of the MPR is in a line-mode test (LMT), which is an important new feature for high-density ULSI DRAM's [11] and will be explained in Section III. In Section IV, the memory cell structure and a scaled-down isolation technique are discussed. Finally, the performance and other features of the RAM are summarized.

II. Array Architecture

A. Twisted-Bit-Line Technique

A reasonable chip size containing 16-Mbit memory cells drives the memory cell to a 4–5 μm^2 size, and the gate length and the isolation width to 0.6 μm. The key features for an array architecture using scaled-down memory cells are: 1) an array noise reduction technique, and 2) a multi-divided bit-line structure which produces sufficient signal voltage and allows a wide operating margin at a 3.3-V power supply. Although the folded bit-line structure offers some array noise reduction, the inter-bit-line coupling noise [9], [12], which is a serious problem in 16-Mbit DRAM's, cannot be cancelled with the folded bit-line structure.

By measurement and simulation, the inter-bit-line coupling capacitance amounts to 21 percent of the total bit-line capacitance, even with a thin tungsten bit line. The signal loss in the worst-case data pattern is more than 30 percent of the total signal amplitude in the folded bit-line structure [9].

In order to suppress the inter-bit-line coupling noise, a TBL array structure is used. Each scaled-down bit line has 128 cells and realizes a small C_b/C_s ratio of six. The signal loss is reduced by 16 percent of the total signal amplitude. Fig. 1 shows part of a 2-Mbit block along with the sense amplifiers associated with the sub I/O bus lines. This 2-Mbit block is divided into eight subarrays by the placement of nine alternate shared (ALS) sense amplifier bands. The RAM operates with a partial activation technique. Only one subarray and a single double-sided ALS sense amplifier of that subarray are activated in a 2-Mbit block.

Manuscript received April 7, 1989; revised June 5, 1989.
The authors are with the LSI Research and Development Laboratory, Mitsubishi Electric Corporation, 4-1 Mizuhara, Itami 664, Japan.
IEEE Log Number 8929847.

0018-9200/89/1000-1184$01.00 ©1989 IEEE

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

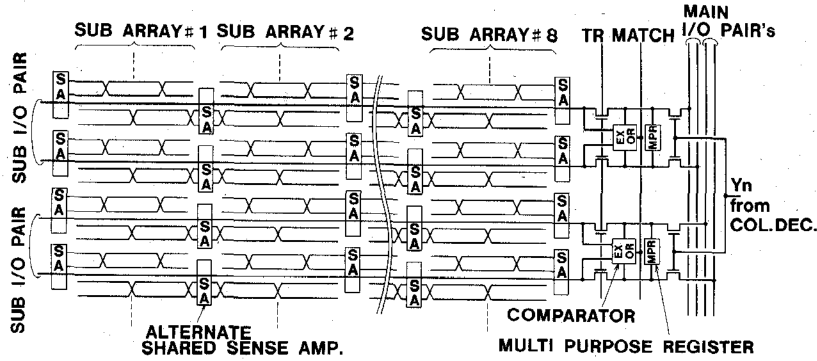

Fig. 1. Part of 2-Mbit block with sense amplifiers associated with sub I/O lines.

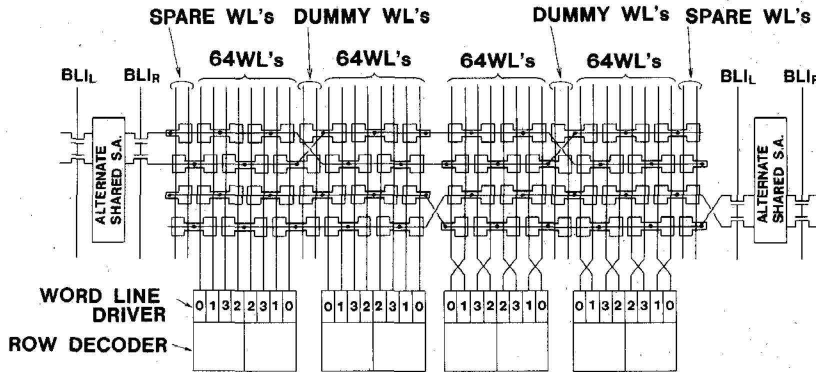

Fig. 2. Physical cell arrangement with dummy and spare elements in a basic subarray.

TABLE I
COMBINATION OF WORD LINE AND DUMMY WORD LINE IN TBL

Word-Line	Dummy Word-Line
A Block 0 , 2 C Block 0 , 2	D1
A Block 1 , 3 C Block 1 , 3	D2
B Block 0 , 2 D Block 0 , 2	D3
B Block 1 , 3 D Block 1 , 3	D4

Only one of the four bit-line pairs is connected to the sub I/O pair. The array charging current is reduced by 1/8. This ALS sense amplifier gives two benefits with little area increase: 1) faster sensing by halving the load capacitance of the sense drive-line, and 2) relaxed layout pitch.

The hierarchical data bus consists of sub and main I/O lines, and second-aluminum sub I/O bus lines, which are arranged so as to connect every four bit-line pairs. The load capacitance value of the sub I/O line is estimated to be less than 0.6 pF. In addition to the reduction and distribution of the load capacitance, an intermediate buffer called the multipurpose register (MPR) is inserted between the sub and main I/O lines to realize high-speed array operation.

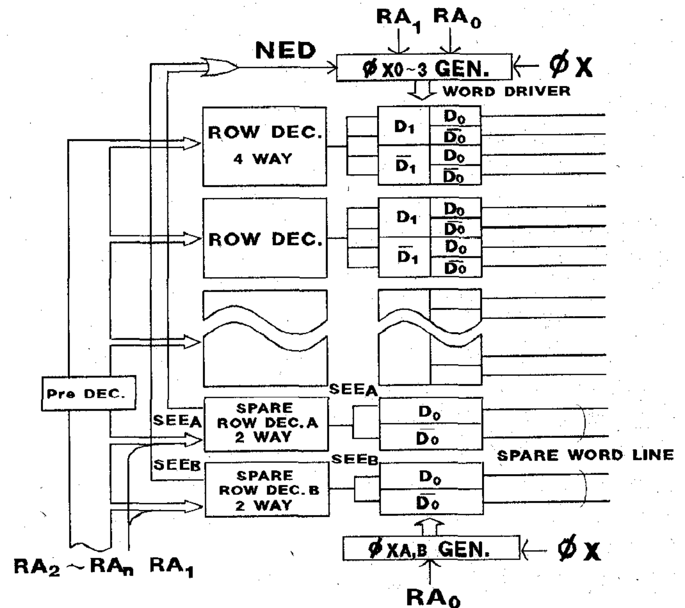

Fig. 3. Block diagram of spare row scheme.

Fig. 2 shows the physical arrangement of cells with dummy and spare word lines in a basic subarray. In this case, the situation fits the dummy reversal technique [13]. The TBL array causes the chip area increase and the complex dummy decoding scheme. To solve these problems, the dummy cells are located at the bit-line twisting intersection and the partly twisted word-line interconnec-

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

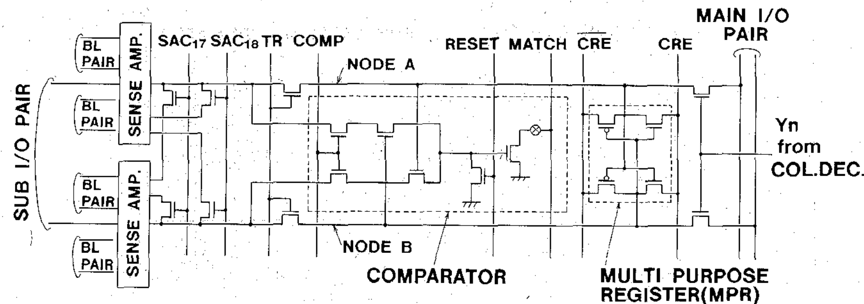

Fig. 4. Multipurpose register and comparator.

tion techniques are introduced. Four sets of dummy cells located at the bit-line twisting intersections allow for regularity of cell arrangement except for the center of the basic subarray. Only the center twist wastes area in the basic subarray. Therefore, the chip size increase caused by the additional dummy and twisting areas is only 0.1 percent.

The partly twisted word-line interconnection simplifies the dummy word-line decoding. The word-line decoding scheme adopts the four-way configuration which consists of a row decoder and four word drivers. One of four word-line drivers is selected by the lower two bits row addresses RA_0 and RA_1. A four-way dummy decoding as well as the row decoding scheme can be realized by the partly twisted word-line scheme (*C* block and *D* block). Table I shows the combination of the word line and dummy word line in TBL. It means that the dummy word line in *A* block and *C* block (*B* block and *D* block) is decoded by the same decoding scheme. For example, whether in *A* block or *C* block, when the even word line is activated the dummy word line D_1 is selected. If the partly twisted word-line layout is not adopted, the dummy word-line decoding would be eight-way decoding.

To obtain a reasonable production yield of 16-Mbit DRAM's with very small physical dimensions and a large chip size, a flexible spare row scheme is introduced. In adopting additional two-way spare word-line driving clocks decoded by only row A_0 (RA_0) address input, one set of four spare word lines is divided into two sets of two spare word lines and each pair of spare word lines can replace any pair of neighboring word lines in a basic subarray as shown in Fig. 3. A consistent dummy word-line selection signal is generated in collaboration with word-line replacement. Therefore, a comparatively large defect, such as two to four word-lines' short circuit, can be repaired, to say nothing of 1-bit failure and word-line snapping.

B. Multipurpose Register (MPR)

Fig. 4 shows an MPR and a comparator circuit. The comparator is provided to perform the LMT. It is an EXCLUSIVE-OR circuit similar to a content addressable memory (CAM) cell, and is comprised of a coincidence detector with a wired-OR match line. The MPR consisting of a full CMOS latch is inserted between the main and sub I/O pairs and is located at both sides of the column

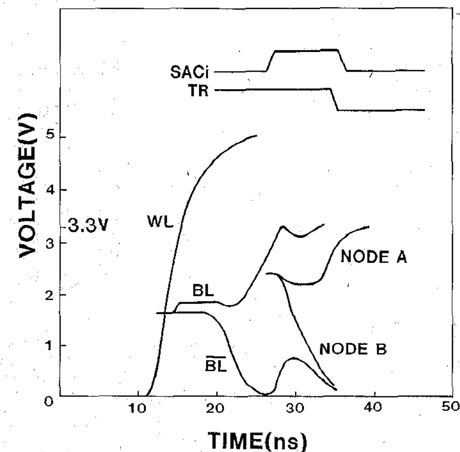

Fig. 5. Simulated waveforms of readout operation.

decoder as explained previously. The functions of the MPR are: 1) intermediate amplifier in a normal cycle; 2) high-speed data register in fast page and static column mode; 3) cache register independent of DRAM array operation; 4) buffer register in copy write mode; and 5) expected test data latch utilized for parallel comparison in LMT.

The MPR operation as an intermediate amplifier in a normal cycle is explained by using the simulated waveforms shown in Fig. 5. At first, the bit-line isolate signal switches the bit lines connected to the ALS sense amplifier. When the word line rises, the information stored in the memory cells is read out on the bit-line pairs. Then, the activated ALS sense amplifiers, which are placed at both edges of the selected subarray, detect and amplify the bit-line voltage level. After sensing the cell data, the readout data are transferred from the ALS sense amplifier to the sub and main I/O bus lines by *SACi* (sense amplifier connecting signal generated from column addresses) and Y_n (column decoder output), sequentially. The driving clocks *CRE* and $\overline{CRE}$ activate the MPR. The activation of the MPR accelerates the amplification of the main I/O bus lines because the drivability of the MPR is higher than that of the sense amplifier. Moreover, the driving of the MPR starts automatically via the ALS sense amplifier before *CRE* and $\overline{CRE}$ are activated. Therefore, the overhead caused by the MPR control sequence is negligible. At the same time as *CRE* and $\overline{CRE}$, the MPR is isolated from

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

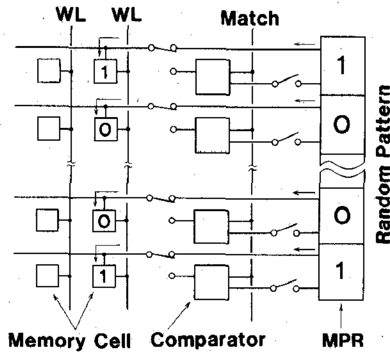

Fig. 6. Principle of copy write in LMT.

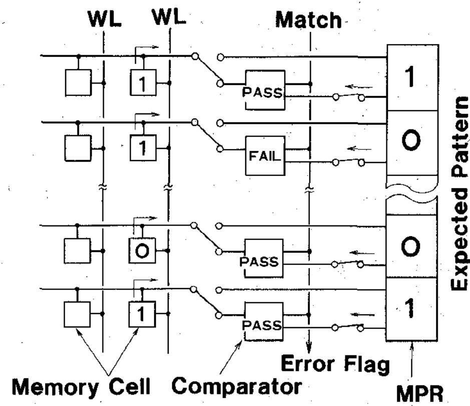

Fig. 7. Principle of parallel compare called line read.

the sub I/O bus lines by *TR*, and the sub I/O lines are isolated from the ALS sense amplifier by *SACi*, sequentially. This isolation sequence improves the data transfer speed because of the reduced load capacitance of the data bus, and also contributes to the reduction of the charging current of sub I/O lines.

III. Line-Mode Test (LMT)

A serious problem with VLSI memory manufacturing is the longer test times required which result in higher chip costs. In order to reduce the test time, the multibit test mode was first implemented in 1-Mbit DRAM [14] and was standardized in 4-Mbit DRAM. The multibit test mode reduces the test time to $1/4 \sim 1/8$ by testing four to eight bits simultaneously, but this multibit test method is limited to no more than 16 bits because of the larger area penalty caused by the additional circuit elements. To overcome this limitation, the concept of parallel testing is expanded to the LMT which simultaneously tests memory cells connected to a word line. The LMT reduces the test time drastically and the MPR enables a random pattern test along the word line.

Fig. 6 shows the principle of a copy write in the LMT. Proceeding the copy write, test data should be written into the MPR utilizing the cache register function which is independent of the DRAM array operation. The MPR permits a random test pattern on a line. This random data pattern in the MPR is written into the memory cells on the selected word-line simultaneously. In the circuit diagram of Fig. 4, these operations are controlled by raising *TR* and selecting *SACi* "high" with *COMP* (LMT compare trigger signal) remaining "low."

Fig. 7 shows the principle of the parallel compare called a "line read." the readout cell data on a word line are transferred through the sub I/O lines. The expected data pattern latched in the MPR is transferred as well. When the test–compare–trigger signal *COMP* goes "high" and *TR* remains "low," parallel-compare is carried out by the comparator, which detects the coincidence between the readout data and the expected data. Then a merged test result appears on the match line. If there exists at least one comparator with disagreement, a fault signal (ERROR FLAG) is sent via the match line to output *DQ*. Thus any fails on a line can be detected in one READ cycle.

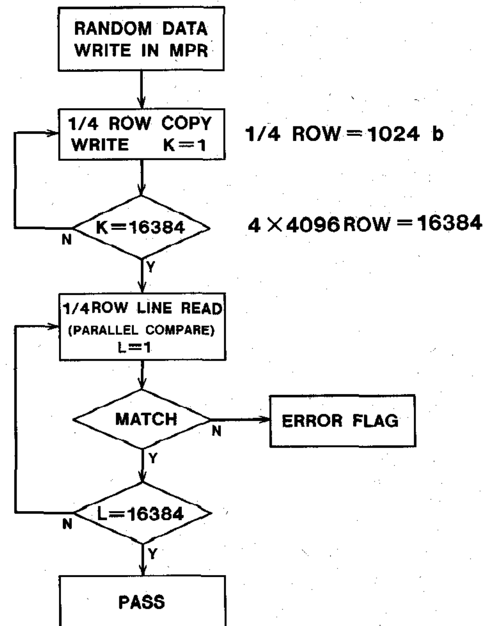

Fig. 8. Flowchart of LMT.

Fig. 8 indicates a flowchart of the LMT in the case of 16-Mbit DRAM. First, random test data are set in the MPR. Then, they are simultaneously written in a fourth of the cells (1024 bits) on a selected word line in one cycle (copy write). 1/4 row means that the number of MPR's is one fourth the bit-line pairs, as shown in Fig. 1. Writing of the random test data in all cells is accomplished in 16K cycles. After that, testing of all memory cells (line read operation) is completed in 16K READ cycles and 16K WRITE cycles. Contrary to the prior-art LMT with only limited test patterns (all 0 or 1 patterns only), this test using the MPR allows the flexible test pattern. It can test with better characterization patterns such as row stripe, column stripe, and checker patterns. These LMT patterns can reduce the conventional chip test time to about 1/1000. An efficient layout of the MPR and the comparator, which is made possible by using a hierarchical data-bus structure, makes a chip overhead of no more than 0.5 percent.

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

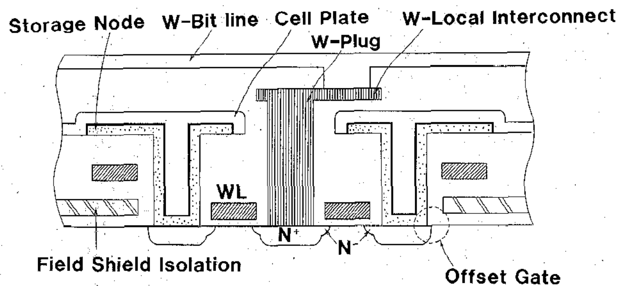

Fig. 9. Cross-sectional view of T-shaped stacked capacitor cell.

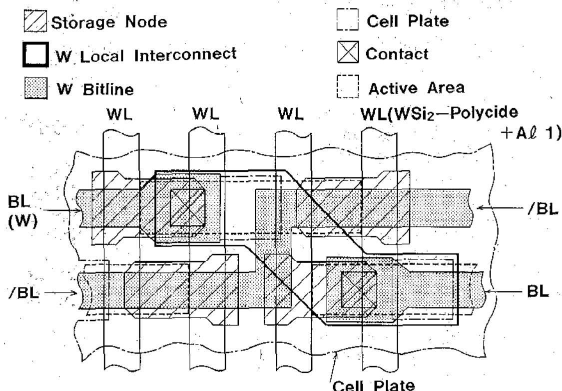

Fig. 10. Top view of stacked capacitor cells at bit-line twisting part.

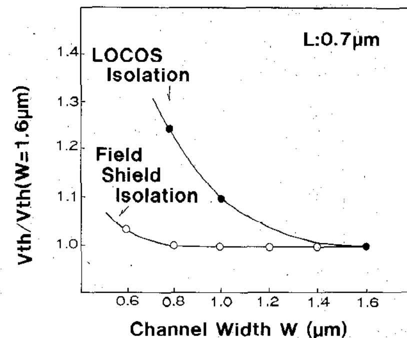

Fig. 11. Narrow channel characteristic of NMOSFET's.

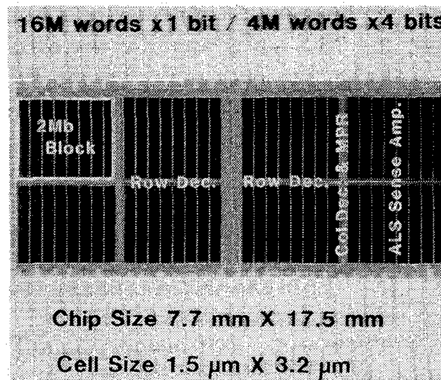

Fig. 12. Photomicrograph of 16-Mbit DRAM.

IV. Device Technology

A cross-sectional view of the memory cells, called T-shaped stacked capacitor cells [15], is shown in Fig. 9. The contact hole connecting the storage node to the N^+ region is formed by a fully self-aligned technique utilizing the sidewall spacer with a word line and field shield plate. The thick CVD oxide film is formed before the storage node process. Using this thick film, a T-shaped storage capacitance of 35 fF was obtained with a 6-nm oxide equivalent dielectric layer. Moreover, the thick CVD oxide film results in a planarized structure and makes the patterning of the storage node easier. The selectively formed CVD tungsten plug in the deep contact hole, and the tungsten local interconnect layer, connect the bit line to the N^+ area. The active areas are isolated by the field shield plate of polysilicon. The field shield isolation transistor has the self-aligned sidewall of an LDD transistor, but it has no N^- region. Thus the field shield transistor has an offset gate structure. This structure reduces the parasitic capacitance between the field shield plate and the source/drain, and has a high threshold voltage of 4.0 V. Fig. 10 shows a top view of the stacked capacitor cells at the bit-line twisting section. This local interconnect wiring is used for TBL's as shown, and minimizes the additional area penalty inherent in the TBL array. Fig. 11 compares the narrow channel characteristic of the NMOSFET's isolated by the field shield with that of those isolated by LOCOS. The horizontal axis is the designed channel width, and the vertical axis is the variation of the threshold voltage which is normalized at a 1.6-μm channel width. The threshold voltage of the LOCOS isolation FET increases rapidly at a channel width under 1 μm because of the severe narrow channel effect. In the case of the field shield isolation FET, the threshold voltage has been controlled to within 10 percent, even for a 0.6-μm channel width device. This is because the field shield isolation is free from the bird's beak and does not need a boron implantation channel stop. Therefore, the field shield isolation has the advantage of eliminating the narrow channel effect, and can be applied to a scaled-down 16-Mbit DRAM memory cell with 0.8-μm dimensions. The gate length of the memory cell transistor is 0.7 μm and the oxide thickness of the active transistor is 15 nm.

In peripheral circuits, the gate lengths of 0.6-μm NMOS and 0.8-μm PMOS transistors with fully self-aligned source and drain contacts [15] provide the same current drivability as that of 0.8-μm (NMOS)/1.0-μm (PMOS) conventional LDD transistors at 5-V operation, respectively. Therefore, these transistor performances and the scaled-down load capacitance give a faster operating speed at 3.3 V than at 5.0 V with an on-chip limiter.

V. RAM Characteristics

A 16-Mbit CMOS DRAM was designed and evaluated. The RAM has been fabricated using a 0.6-μm twin-well CMOS process and double-level aluminum wiring. Fig. 12 shows a photomicrograph of the RAM. The memory array is divided into eight 2-Mbit blocks by the row decoders

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | sales@docketalarm.com | 1-866-77-FASTCASE

fastcase®
Smarter legal research.