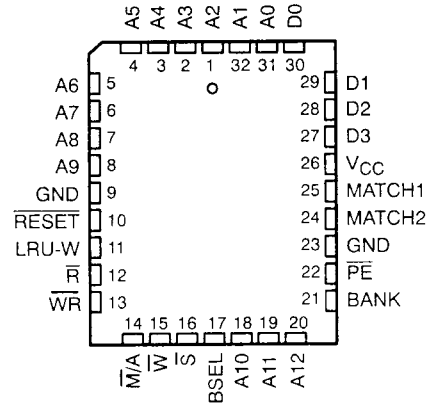
D3365, JANUARY 1990–REVISED JUNE 1990

- **Address to Match Time . . . 17 ns Max**
- **2-Way Architecture Significantly Improves Hit Rate**
- **Implements LRU Replacement Allgorithm**
- **Useful for Bus Watching**
- **On-Chip Parity Generator and Checker**
- **Easily Expandable in Depth and Width**
- **Reliable Advanced CMOS Technology**
- **Fully TTL Compatible**

FM PACKAGE
(TOP VIEW)



description

The SN74ACT2160 is a valuable building block for implementing fast two-way set associative caches. This device consists of two separate 8K x 5 RAMs for tag and parity storage, an 8K x 1 LRU RAM, two high-speed comparators, and the control circuitry necessary to give the designer the freedom to determine how this device will be used. The SN74ACT2160 also includes single-entry invalidation circuitry and parity generation and checking for ease of design and high system reliability.

The SN74ACT2160 is fabricated using advanced silicon-gate CMOS technology for high speed and simple interface with bipolar TTL circuits. By combining the SN74ACT2160 with programmable logic, a cache can be constructed that specifically addresses the individual system requirements. Significant reductions in cache memory component count, board area, and power dissipation can be achieved by using this device.

direct-mapped versus two-way set associative caches

A cache memory is a small high-speed memory that is used to store a portion of the data found in the larger main memory to achieve optimum processor performance and to reduce main memory bus traffic. Since the cache memory is smaller than the main memory, only part of the address, the least significant bits referred to as the index, is used to address the cache memory. The most significant address bits, called the tag, are stored along with the cache data and are used to identify what data is stored in an indexed location. When the processor requests data, the index portion of the processor address points to a word of data in the cache-data RAM and to a tag in the cache tag RAM. If the upper portion of the processor address is equal to the stored tag, a cache hit is said to occur and the cached data can be immediately sent to the processor.

In a direct-mapped or one-way set associative cache, only one data word and tag exist in cache for each index. This means that when the processor requests data, only one cache memory location can contain the requested data. Also, if the requested data is not in the cache and the cache is updated, the data in the indexed cache memory location will be written over regardless of how recently it has been used.

In a two-way set associative cache, two data words and tags exist for each index. This means that the requested data can reside in one of two cache locations. When a miss occurs and the cache is updated, the least recently used data can be written over. As would be expected, studies have shown that the hit rate improves significantly when using a two-way cache design over a one-way or direct-mapped cache design. Through use of the 'ACT2160, the logic complexity and parts count usually associated with a two-way cache are greatly reduced.

address comparison

The 'ACT2160 compares the contents of the memory location addressed by A0-A12 with the address bits (or tag) applied at D0-D3. An equality is indicated by a high level on the MATCH1 or MATCH2 outputs. MATCH1 is high when the applied tag is equal to the stored tag in bank 1. MATCH2 is high when the applied tag is equal to the stored tag in bank 2.

writing to the cache tag RAMs

The manual/auto ($\overline{M}$/A) input on the 'ACT2160 provides two methods of selecting which tag bank will be written to when the write input ($\overline{W}$) is taken low. When $\overline{M}$/A is low, the bank select input (BSEL) selects the bank to be written to. BSEL low selects bank 1 and BSEL high selects bank 2. When $\overline{M}$/A is high, the least recently used (LRU) circuitry automatically selects the bank written to when $\overline{W}$ is taken low.

writing to the cache data RAMs

When a read or a write miss occurs and the cache is updated, the BANK output will indicate which bank the data should be written to. If BANK is low, bank 1 should be written to. If BANK is high, bank 2 should be written to. When writing a tag with $\overline{M}$/A low, the BANK output will not indicate which bank is being selected by the BSEL input. BANK is the output of the internal 8K x 1 LRU RAM. When a write hit occurs, the match outputs will indicate which data bank to write to.

LRU replacement circuitry

A concept commonly referred to in cache design is the property of locality. An aspect of the property of locality says that the information currently in use is likely to be used again soon. Based on this property, it is desirable to replace the information that has not been used recently when writing to cache. With a set size of two, this is easily done using one bit to indicate which of the two addressed locations is oldest or least recently used. The 'ACT2160 contains an 8K x 1 RAM and the necessary circuitry to implement the LRU replacement algorithm.

The $\overline{M}$/A input allows the user to choose between automatic LRU and manual replacement. When $\overline{M}$/A is high, the LRU RAM output selects which bank to write to. When the LRU bit for a given address is low, a write pulse will write D0-D3 to bank 1. When the LRU bit for a given address is high, a write pulse will write D0-D3 to bank 2. The LRU RAM is updated every time a write, a match (with LRU-W signal), or a word reset occurs.

When a write occurs with $\overline{M}$/A high, the addressed LRU bit is inverted and written back in so that the next write with $\overline{M}$/A high to that address will be to the other bank. When a write occurs with $\overline{M}$/A low, the bank is selected by the BSEL input and the addressed LRU bit is adjusted so that the next write to the same address with $\overline{M}$/A high will be to the other bank.

With a match output high indicating a match, the LRU RAM will also be updated when the LRU-W input is taken low. The logic level at each match output is fed back internally to the LRU circuitry. If MATCH1 or MATCH2 are high, the LRU-W input provides an LRU write timing signal that causes an internal LRU write pulse to be generated. With MATCH1 high, the LRU bit is set high so the next write to the same address with $\overline{M}$/A high will be to bank 2. With MATCH2 high, the LRU bit is set low so the next write to the same address with $\overline{M}$/A high will be to bank 1. When cascading these devices for wider address coverage, the MATCH1 outputs must be wire-ANDed together so an LRU write will not occur unless all MATCH1 outputs are high. In the same manner, the MATCH2 outputs (in width) must be tied together. When MATCH1 and MATCH2 are forced high during deselect, write, read, word reset, or reset, and LRU-W is taken low, a false LRU write will not occur.

for complete data sheet

The complete version of this data sheet and application information can be found in the *Cache Memory Management Data Book*, Literature #SCAD002. To obtain a copy of this data book, contact your local TI sales representative or call the TI Customer Response Center at 1-800-223-3200.