SP 22.2: 64kB Sum-Addressed-Memory Cache with 1.6ns Cycle and 2.6ns Latency

R. Heald, K. Shin, V. Reddy, I-F. Kao, M. Khan[1],
W. Lynch, G. Lauterbach, J. Petolino

Sun Microelectronics, Palo Alto, CA
[1]Texas Instruments, Stafford, TX

This circuit combines a sum-addressed-memory (SAM) cache with delayed-reset logic circuitry, enabling cache access with a two-cycle-latency for a 600MHz third-generation superscalar processor implementing the Sparc V9 64b architecture [1-5].

The SAM decoder uses extensive parallelism of conventional RAM decoders, together with the fact that unlike an adder, an (A + B = K) equality test requires no time-consuming carry propagation [6]. A SRAM decoder can be imagined to perform a large AND function of address lines, where these AND gates test for a sum, S, equal to the constant, K, for each word line K. To perform this parallel test, the address sent to the final decoder cannot depend on K. This requires some added logic within the predecoder to precondition the addresses prior to the final decoder. The preconditioning can be a bit at a time, or in groups of bits as in the two-bit SAM predecoder logic shown in Figure 1. The predecoder uses groups of 3 base and 3 offset bits to produce 2 out of 8 predecoded signals for the final decoder. Two-hot encoding allows parallel evaluation of the carry-in. Unique word line addressing prior to carry-in determination is by dividing the array into segments such as odd and even lines.

For a 64kB 4-way SAM cache with 512 256b lines, the selected line is addressed with the SAM row decoder shown in Figure 2. The column is selected by the 1 of 4 column decoder shown in Figure 3. The odd or even array is determined by the array decoder shown in Figure 4. Address bits 13 through 5 are used for addressing the word lines, and bits 4 and 3 select the double word within the line. Bits 4 through 0 generate the carry_out_4 for the odd/even select and bits 2 through 0 generate the carry_out_2 for the column decoder. The column decoder shown logically in Figure 3 combines the carry_out_2, c2, from these low-order bits with bits 3 and 4 to provide a unique one-of-4 double word selection. Column decoder timing is less critical than row decoder timing, so the carry circuit does not limit access time. The array decoder in Figure 4 produces the array choice in time to power the correct sense amplifiers. The bank select arrives in time to drive a normal 4 to 1 multiplexer between the sense amplifiers and the block data output drivers.

For highest speed with a limited power budget, delayed-reset logic realizes this SAM cache. Delayed-reset logic is similar to post-charge logic and self-resetting logic used in other high speed RAM designs [2,3,4]. Extension of these techniques to include the SAM decoder requires only revised predecoder and column decoder circuits. Delayed-reset logic (and most other cousins of domino logic) work best in high fan-out situations. The chosen predecoder realization in Figure 5 combines the 2b XOR and input AND/OR logic into a nMOS stack. The output AND is done in the P-stage drives the predecoder outputs to the final decoder. Some keeper and reset devices are omitted from this schematic for clarity.

Compared to the common 1:8 tree predecoder, the SAM decoder increases block delay by about 110ps, or about one loaded gate delay. As the SAM predecoder is a 2:8 output, the final row decoder, while conventional in design, has a slightly longer delay due to the additional input. Thus, a SAM decoder gives a cache access of less than 2 delays greater than a normal cache while eliminating a separate address add stage prior to cache access.

The SAM decoder circuitry drives an array of six-transistor memory cells organized in 16 blocks of 256b by 128w. The read and write column decoders select the four double words for output or input as needed. The sense amplifier is a low power CMOS latch feeding a 4-way hit-multiplex circuit. The hit-multiplex output is fed through an alignment-multiplex and exits the SAM cache block. A micro-tag containing 8b of the complete tag is included in the SAM cache block and is SAM accessed in parallel with the cache. Its small size allows the access and hit compare to take place in the time the cache is accessed and the data amplified by the sense amplifier. The hit signal is generated in time to control the 4-way multiplex circuit. Redundancy is provided by two extra word lines and two spare bit lines in each 4k block. The RAM latency is controlled without flip-flops in the memory cell access path by adding latency-control latches to the output buffer.

A bit slice of the RAM is fabricated in a 0.25μm six-metal process and tested to check data flow and timing. The decoder and output structure are simplified to fit the test environment. The test structure critical path consists of: address-input flip-flop, delayed-reset non-SAM predecoder, delayed-reset decoder, word line, memory cell, bit line, sense amplifier, and output latency control. The path operates at 720MHz at 1.6V and 100°C. This compares well with the simulated speed of 760MHz.

The IDS waveforms in Figure 6 show a write operation followed by one read operation and a portion of a second read. The second positive-going clock begins the read access, resulting in the positive pulse of the selected word line, WL, about 600ps later. The negative-going global sense amplifier strobe, SAE_L, about 700ps following the clock, triggers the sense amplifiers through local inverters. The data output, ICA_DO[0], appears long after SAE_L due to latency control circuitry. In the 64k cache, there is more delay between the clock and array access due to wiring. The cache output delay is not limited by the latency circuit while high-data throughput corresponding to the clock frequency is maintained.

Acknowledgments:

The authors acknowledge support of colleagues in Sun Microelectronics SRAM group and thank J. Kaku, K. Tam, L. Rodgers, and S. Mehrotra for help in paper preparation.

References:

[1] Lynch, W., et al., "Low-Latency Cache Indexing: XOR and SAM Caches," Sun Labs Report SMLI 93-0350, 1993.

[2] Lev, L., et al. , "A 64-b Microprocessor with Multimedia Support," IEEE J. of Solid-State Circuits, vol. 30, pp. 1227-1238, Nov., 1995.

[3] Heald, R., J. Holst, "A 6-ns Cycle 256 kb Cache Memory and Memory Management Unit," IEEE J. of Solid-State Circuits, vol. 28, pp. 1078-1083, Nov., 1993.

[4] Chappell, T., et al., "a 2-ns Cycle, 3.8-ns Access 512-kb CMOS ECL SRAM with a Fully Pipelined Architecture," IEEE J. of Solid-State Circuits, vol. 26, pp. 1577-1585, Nov., 1991.

[5] Weaver, D., T. Germond, *The SPARC Architectural Manual*, Version 9, Prentice Hill, 1994.

[6] Cortadella, J., J. Llaberia, "Evaluation of A+B=K Conditions Without Carry Propagation," IEEE Transactions on Computers, vol. 41, pp. 1484-1488, Nov., 1992.

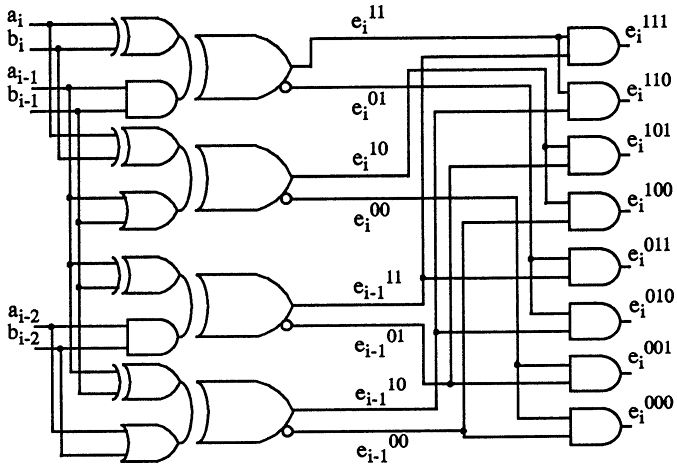

Figure 1: SAM 2b predecode logic.

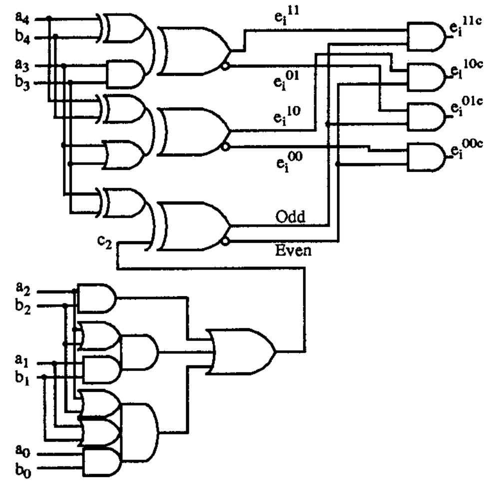

Figure 3: SAM 1 of 4 column decoder.

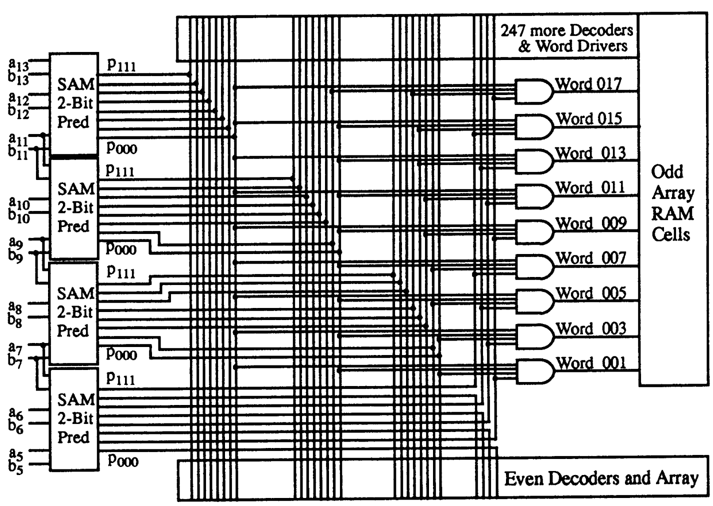

Figure 2: SAM row decoder for 64k cache.

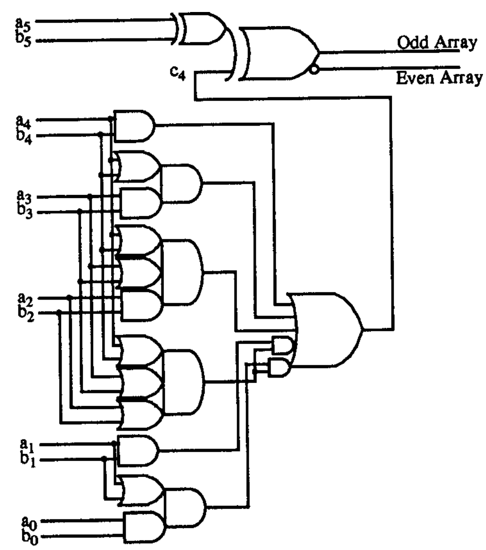

Figure 4: SAM array decoder.

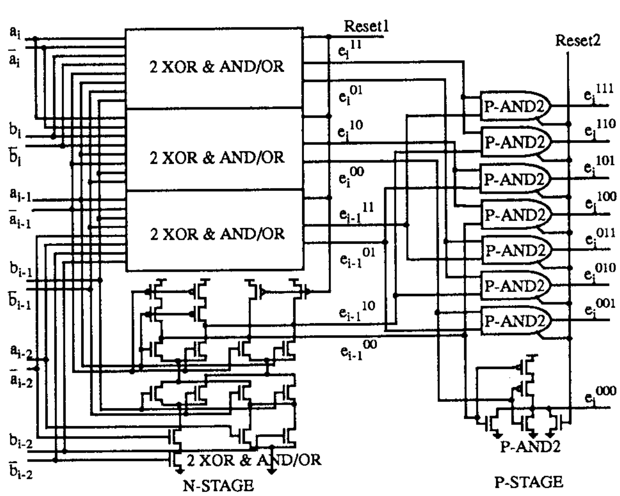

Figure 5: SAM predecoder circuit.

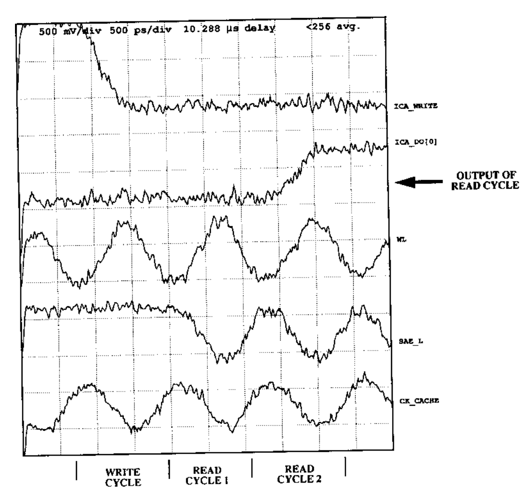

Figure 6: Test circuit waveforms.

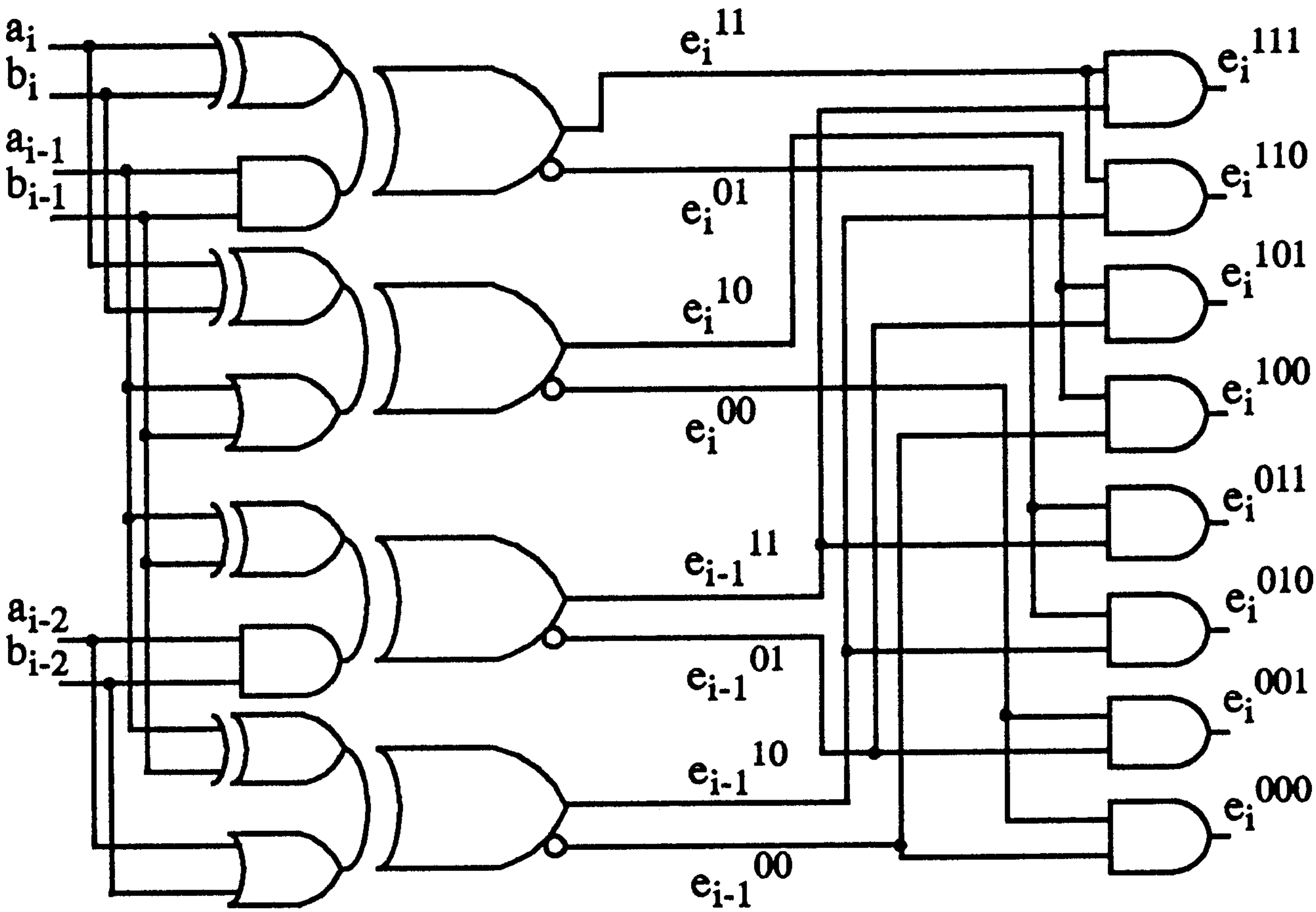

Figure 1: SAM 2b predecode logic.

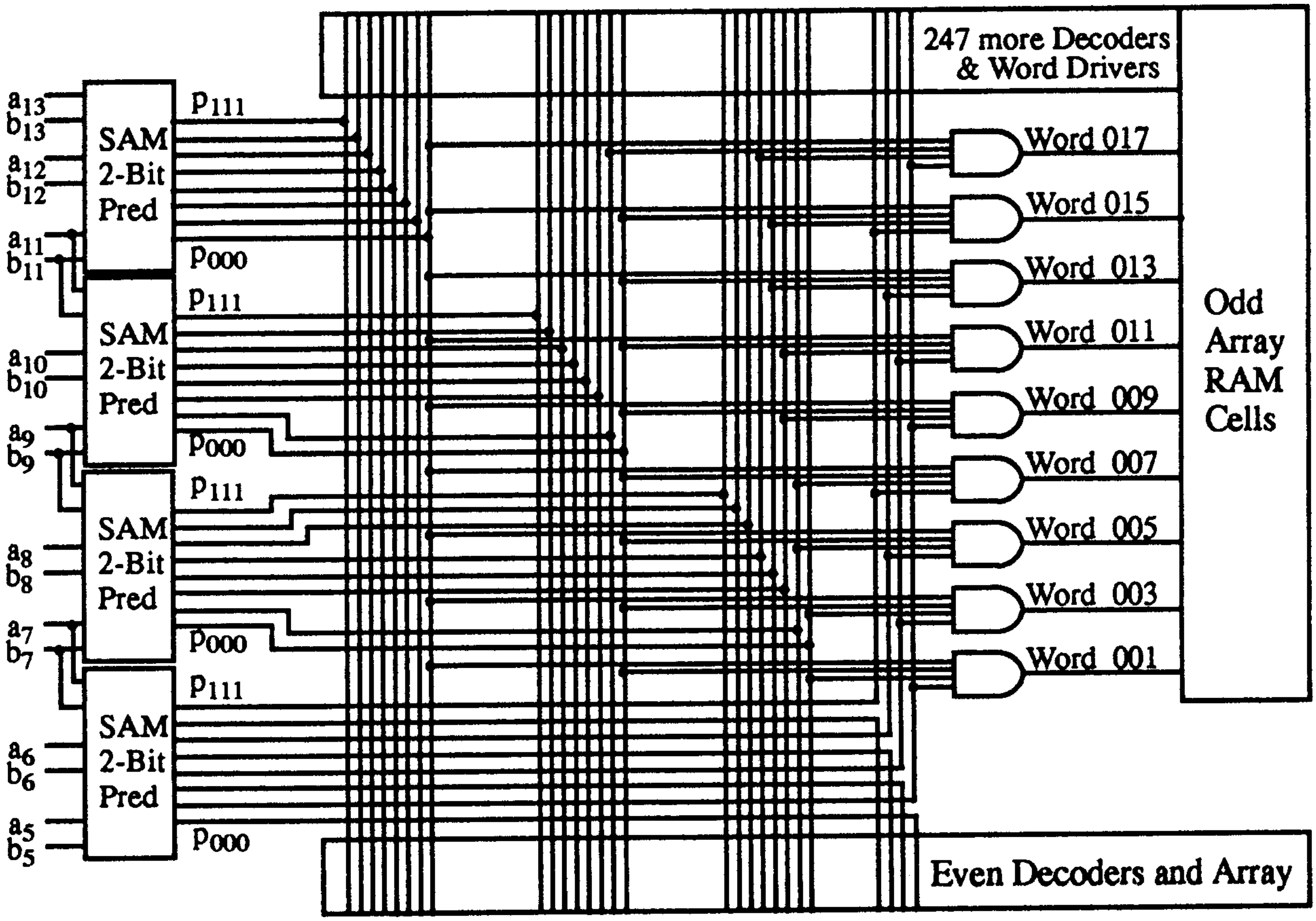

Figure 2: **SAM row decoder for 64k cache.**

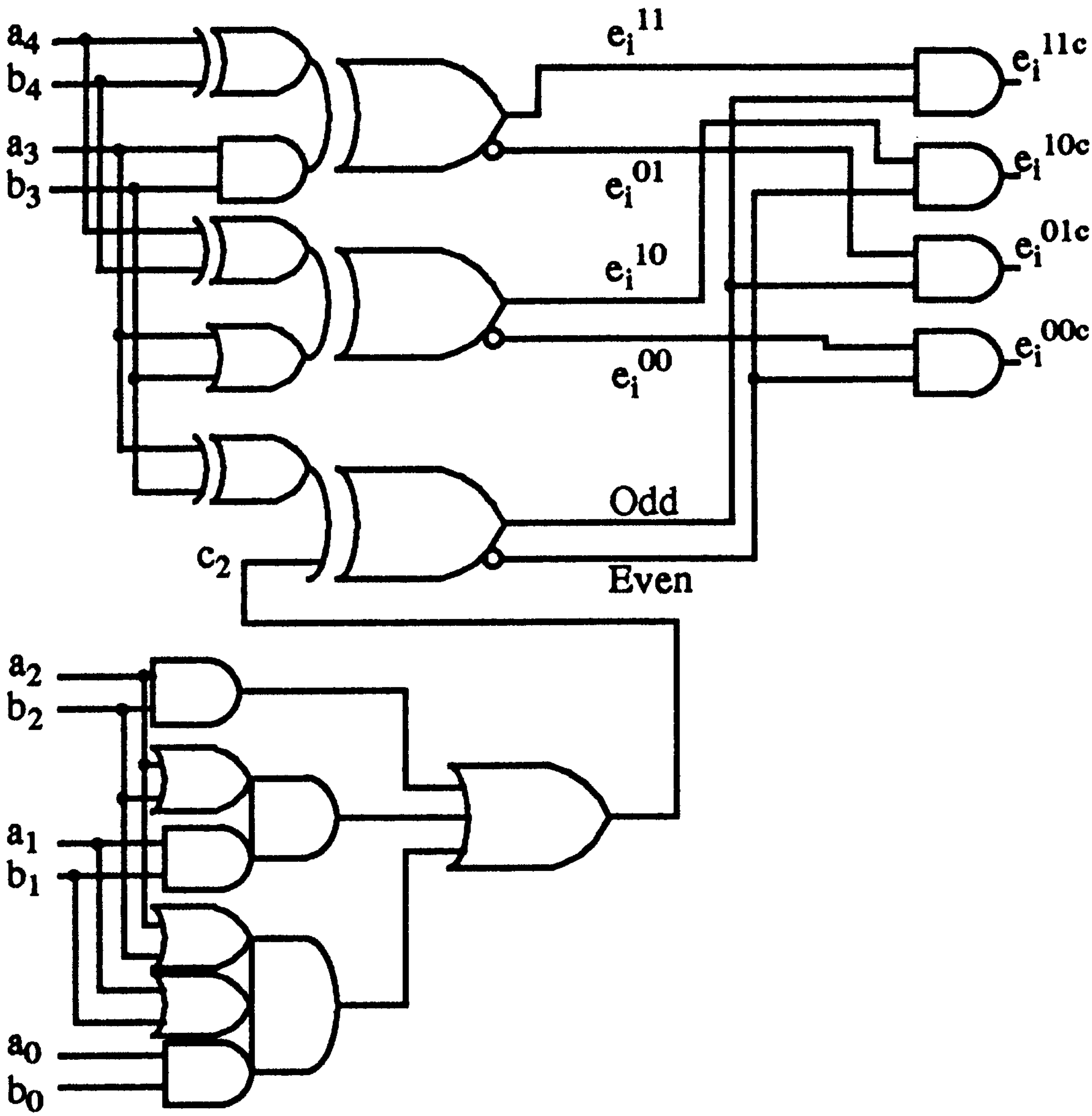

Figure 3: SAM 1 of 4 column decoder.

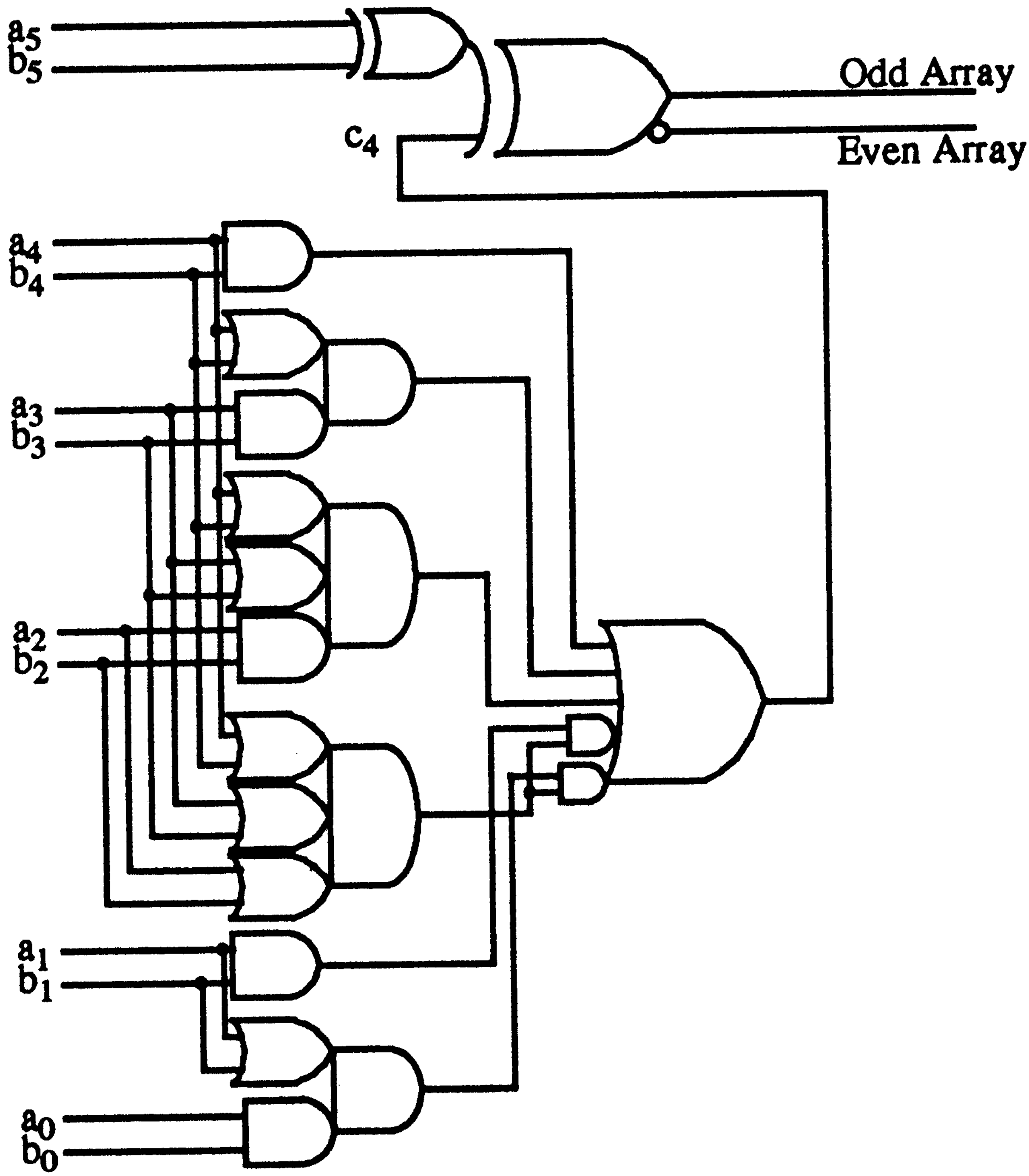

Figure 4: SAM array decoder.

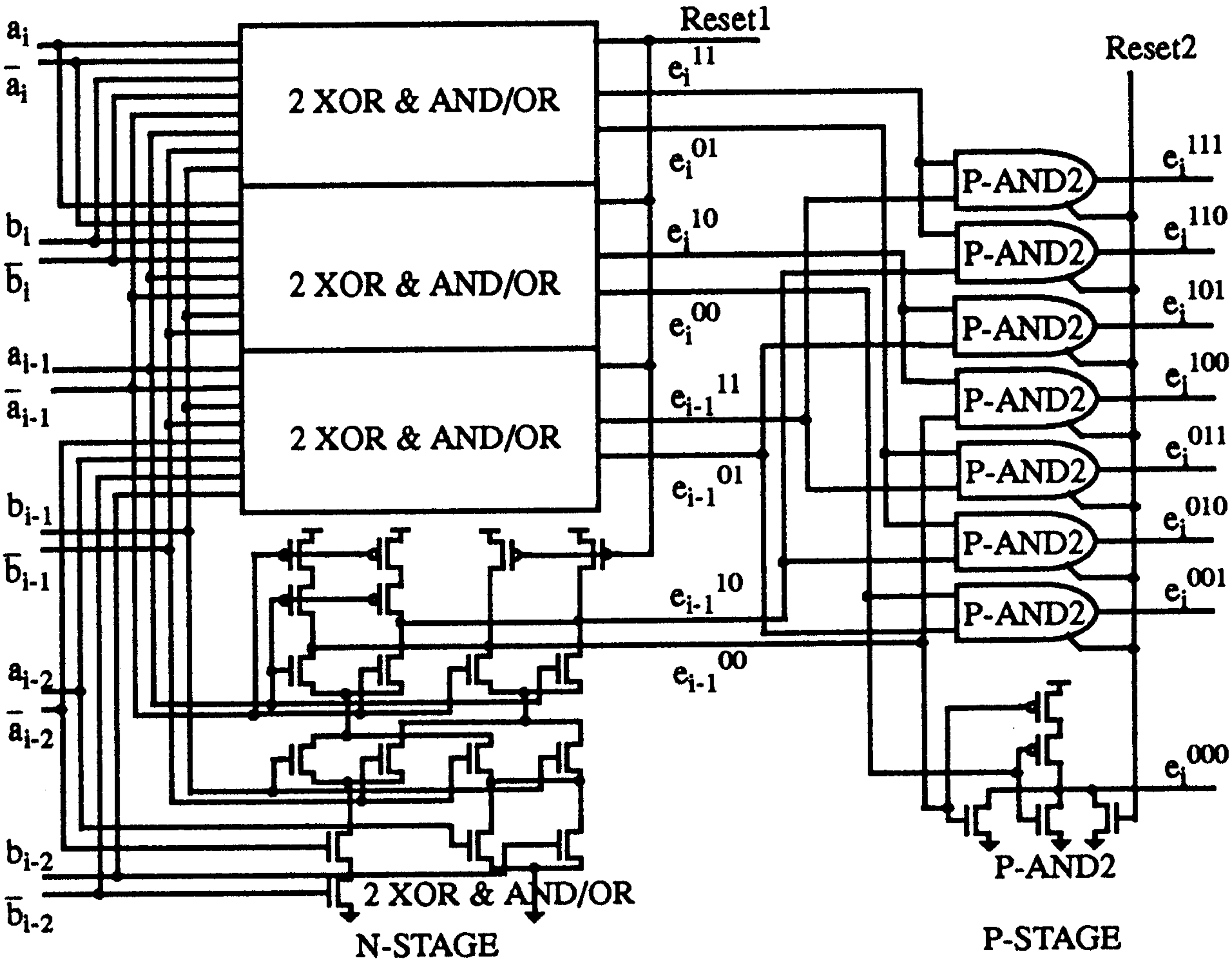

Figure 5: **SAM predecoder circuit.**

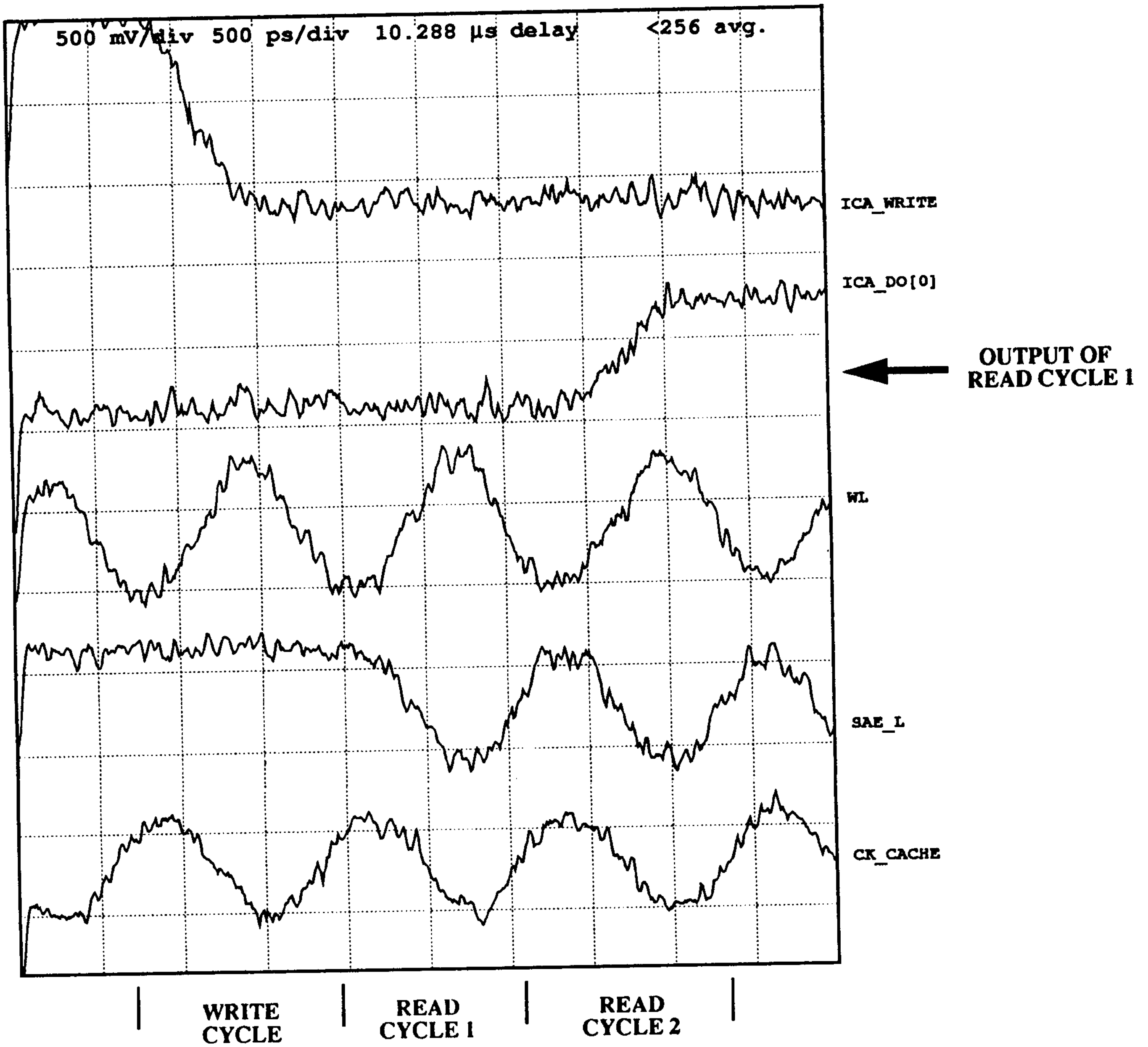

Figure 6: Test circuit waveforms.