A 200 MHz 0.8 μm BiCMOS MODULAR MEMORY FAMILY OF DRAM AND MULTIPORT SRAM

Allan L. Silburt*, Richard S. Phillips†, G.F. Randall Gibson*, Armin G. Bluschke*, Rajesh K. Verma*, Peter M.J. Diedrich*, Stephen P. Kornachuk*, James S. Fujimoto* and Benoit Nadeau-Dostie*

* Bell-Northern Research Ltd.
P.O. Box 3511, Station C
Ottawa, ON, K1Y 4H7, Canada

† Bayview Technologies Inc.
565 Bayview Drive, Box 50
Constance Bay, ON, K0A 3M0, Canada

Abstract: A family of modular memories has been designed in a 0.8 μm BiCMOS process based on a synchronous self-timed architecture. Nominal access and cycle times are 5 ns for 64KBit blocks of 1, 2 and 4 port SRAM as well as a DRAM using a four transistor (4T) core cell.

INTRODUCTION

The goal in developing the modular RAM family for the Northern Telecom Electronics Ltd. 0.8 μm BiCMOS Telecom process (BATMOS) [1], summarized in Table 1, was to design circuitry which could be adapted with few changes to implement SRAM, multiport SRAM and 4T DRAM over the modularity ranges outlined in Table 2. The result is a four member family with virtually identical schematics, and layout optimized to achieve high density and speed. This paper will discuss the modular self-timed specification and design, its extension to multiport and dynamic RAM cells, and the measured performance.

Table 1: BATMOS Process Summary

Vertical NPN:	Emitter size (min)	0.8 × 4.0 μm
	F_t	11 GHz
NMOS and PMOS LDD:	Lmin	0.8 μm
	VTP / VTN	-0.85 / 0.80 V
	tox	17.5 nm
Interconnect:	Metal 1 line 0.8 μm, pitch 2.0 μm	
	Metal 2 line 0.8 μm, pitch 2.0 μm	
	Metal 3 line 1.0 μm, pitch 2.4 μm	
	Stacked via and contact structures	
	Local Interconnect	

Table 2: RAM Family Modularity Range

Parameter	Min	Step	Max
Words	8	1	8K
I/O Bits per Word	1	1	64
Total Bits	8	–	64K
Rows	4	2	256
Columns	2	2	256
Columns per I/O Bit	2	2	32
Row Address Bits	2	1	8
Column Address Bits	1	1	5

SELF-TIMED RAM SPECIFICATION

The specifications for the modular memory family are geared directly for ease of use by the system designer and integration within a synchronous design methodology. A read cycle timing diagram for a generic port is shown in Figure 1.

A self-timed memory architecture is used with a registered CMOS interface on all input pins. This provides the best possible integration of embedded memory into large synchronous ICs since the interface timing is similar to an edge triggered flip-flop. Worst case setup and hold times are 3 ns and 0 ns respectively. This natural interface to a synchronous system also allows the creation of simple bus cycle models for the memory which map easily into high level synthesis tools.

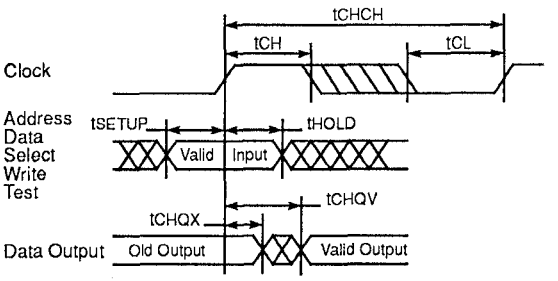

Figure 1: Read Cycle Timing Specification

The buffered outputs of all registers are available to the system designer and can be used, for example, to form counters and scan registers. They also constitute the majority of circuitry required for a Built in Self Test (BIST) circuit [2]. Characterizing the memory block and the input registers with an analog simulator as a single unit and modeling it as such at the logic level, improves the accuracy of the analysis of system timing paths through the RAM. The system designer need not rely on a logic simulator and pre-layout wire load estimates to margin the critical register-to-memory input timing.

The self-timed shut down yields a power dissipation characteristic that is inversely proportional to clock frequency and independent of duty cycle making it well suited to both high and low speed applications. In addition, the length of the active portion of the cycle decreases with faster process conditions. As a result, the power dissipation is less sensitive to process and temperature than other RAMs which depend on level sensitive clock or enable inputs that must be derived from a fixed duty cycle system clock. A synchronous memory select pin is also included to allow system level block subdecoding. Since unselected blocks draw no current, this is useful for trading off power and area at the chip level when constructing large memory functions.

Two test modes are included to provide compatibility with the SCAN and memory BIST test methodologies used at BNR. The first test mode reconfigures the memory data input / output path to a scan compatible mode in which the data inputs bypass the memory core and are transferred directly to the data output pins. This ensures controllability of the memory outputs during scan testing of the logic around the RAM. The

IEEE 1992 CUSTOM INTEGRATED CIRCUITS CONFERENCE

0-7803-0246-X/92 $3.00

registers on all input pins can be included in a scan chain to ensure scan fault coverage right up to the RAM boundary. The second test mode allows a special pattern to be applied to multiport RAMs that checks for shorts between word lines and bit lines from opposite ports as described in [2].

SELF-TIMED ARCHITECTURE

A simplified schematic of the signal and control path for a generic port is shown in Figure 2. A clocked latch style CMOS data path is used to enable fast power-up and shut down of the sense circuitry on each cycle. Sensing schemes requiring steady state bias, as are often used in BiCMOS SRAMs [3,4], are not suitable to this application. BiCMOS buffers are, however, used extensively to drive heavily loaded lines such as the predecoded addresses, word lines and control clocks.

The memory operation is divided into three phases; decode, sense / write and reset. In the decode phase, all inputs are latched, decoding information is propagated and the sense or write circuitry is enabled. In the second phase, bit line data is presented to the sense amplifier / latch or write data is propagated to the bit lines. Finally, the reset phase is enabled by the self-timing path returning the memory to its quiescent state drawing no DC power.

Address Decode

A pulsed, predecoded addressing scheme is used to limit decoding power and design complexity while achieving minimum decode delay over the modularity range. The address inputs are predecoded in pairs and enabled by the system clock. In this manner, a minimum number of predecoded address lines are pulsed in a given cycle – one of four per two address bits as compared to two of four if predecoding were not used. Static decoding of the predecoded address lines selects the active word line and columns while avoiding the requirement for the large decode clock buffers that would be required by a dynamic decoder.

Model Decode Path

The internal self-timing circuitry mimics the address decode operation using the model paths shown in Figure 2. This is essential in achieving tightly margined high speed operation over the full modularity range under all operating conditions. Address predecode buffers identical to the active predecode buffers drive the model predecode signals through the X- and Y-decoder arrays to the model X- and Y-decoders. By accurately loading the model predecode signals with equivalent decoder loads and busing them the full length of the array along with the active predecoded address signals to the model decoders, the transmission line loading effects are simulated. The X-decode and word line delay are simulated by the model word line signal which is generated by the model X-decoder and bused through the memory array. Slightly modified memory cells are used to load the model word line for maximum accuracy. The model column select signal simulates the Y-decode operation using the model Y-decoder.

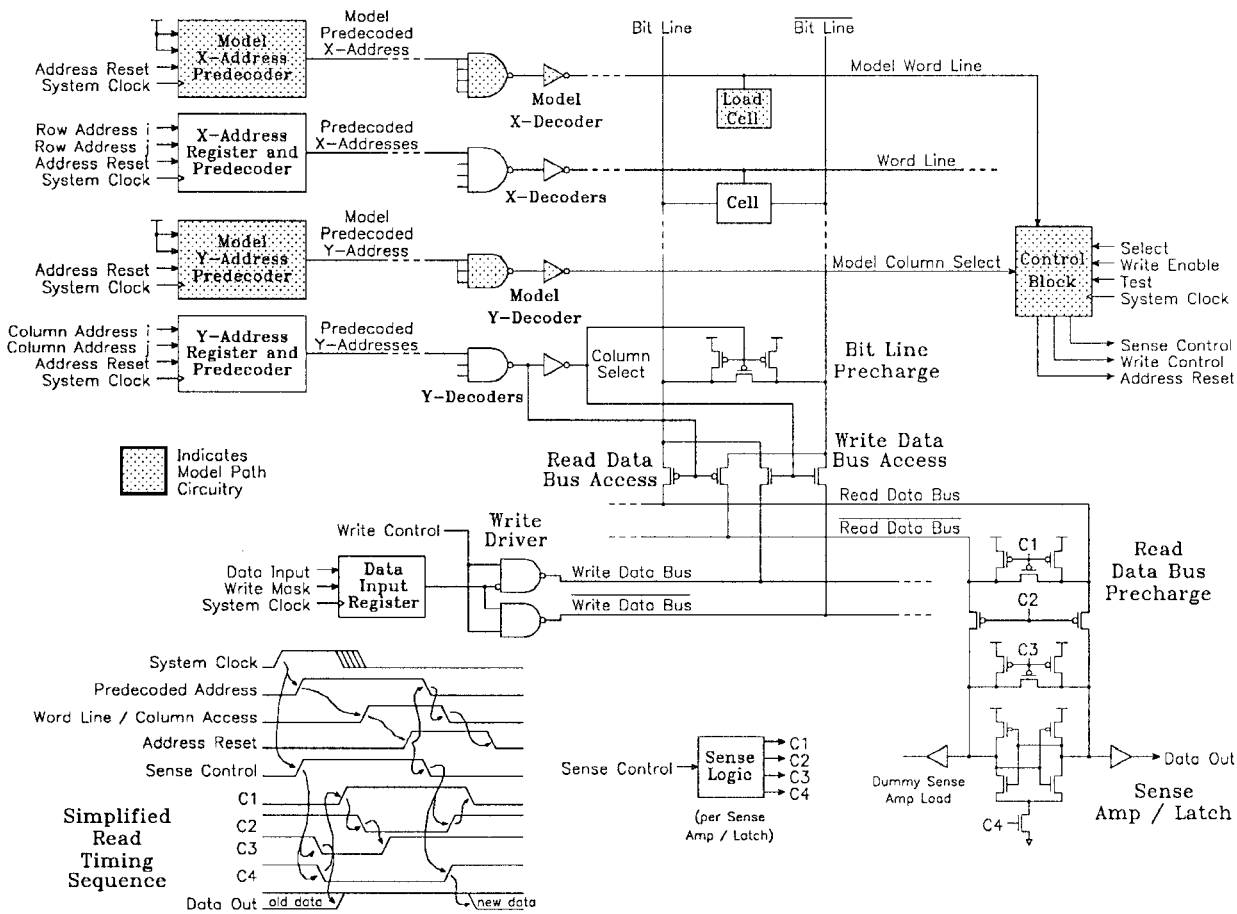

Figure 2: Signal and Control Path Schematic for a Generic Port

Bit Line Access

The column select signal controls bitline precharge as well as the column access function. This simplifies the overall system timing by eliminating the need for global precharge clocks.

Separate read and write data buses are used. This allows the bit line access to be optimized to the different requirements of read and write operations. On reads, the bitline differential is limited to less than 1 V from the VDD precharge level before the word line is deselected by the self-timing loop. The PMOS read data bus access devices pass this signal to the sense amp whose high common mode rejection ratio permits sensing within this narrow band. The NMOS write access devices are off since both write data lines are held high. This isolates the write driver loading from the signal path increasing the signal amplitude. On a write, the NMOS write access devices allow the full 'zero' level to be written to the bit lines. Although on, the PMOS read access devices have negligible impact on write performance and power due to their small size.

If a column is not selected, the read and write access remains off while the bit line precharge devices are on, acting as bit line clamps. Although additional current is consumed through the cell by clamping the bit lines, the elimination of the global control clock which would otherwise be required to switch the precharge network results in lower overall power consumption.

Self-Timed Reset

The self-timed reset of the memory is initiated by the reception of valid model word line and model column select signals at the control block. The address reset signal is generated to reset the predecoded address signals. This deselects the word line and accessed columns which reestablishes the bit line precharge levels. In this manner, the memory is fully deactivated and returned to its quiescent state in which no DC power is consumed while awaiting another cycle.

Sense Amplifier / Latch

A clocked latch style sense amplifier design is used to take full advantage of the synchronous nature of the architecture. Its high common mode rejection makes it ideal for the VDD level bit line precharge.

The sense control signal is enabled by the system clock during reads. It triggers the sense logic associated with each sense amplifier / latch which disables and precharges the sense latch via C4 and the C3 pulse, disables the read data bus precharge via C1, and opens the PMOS data bus isolation devices via C2.

The falling edge of the sense control signal is triggered by the self-timed reset sequence. This self-timing network guarantees sufficient data differential is present on the read data bus prior to sensing. Sensing is initiated from C4 while the data bus is isolated by C2. The read data bus precharge is then reestablished via C1.

The data output is buffered to isolate the sense latch from any external disturbances and provide additional drive. Alternatively, an optional register stage may be added to pipeline the data output. In this manner, the access time (tCHQV) may be reduced to the short 1.0 ns clock-to-q of the flip-flop stage at the expense of one clock cycle latency. Maximum operating cycle time (tCHCH) is not affected.

MULTIPORT IMPLEMENTATION

A symbolic layout for each RAM type and its associated core cell schematic is shown in Figure 3. A core cell area and memory performance comparison for 8K × 8 blocks is shown in Table 3.

Many multiport SRAM designs have dedicated read or write only ports and use single ended cell access schemes [5,6,7]. Although single ended cells may offer some core density improvements when tailored to a specific application, a fully differential cell provides superior read access times for a given

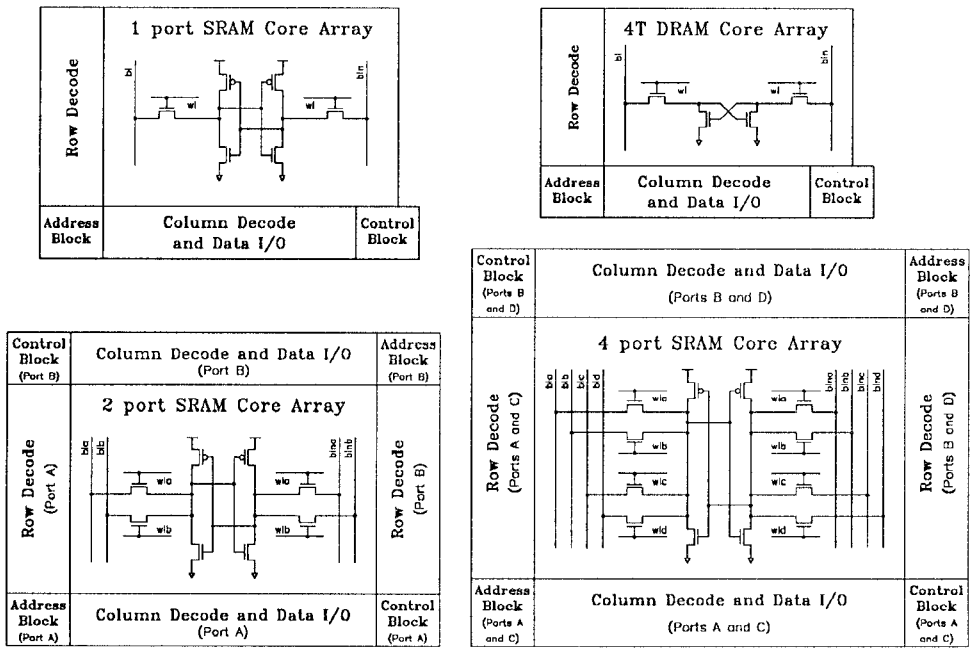

Figure 3: Symbolic Layout Diagram with Core Cell Schematics for Each RAM Type

cell area due to the improved speed attainable with differential sense schemes.

In the architecture described here, an access transistor pair is added per cell to enable fully independent read / write capability for each port. The symmetry this imparts to a multiport memory yields simplicity in design, layout, simulation and built in self-test circuitry. Each port functions independently with its own control block and model signal paths yielding a highly extensible design with broad system applications. Each core cell is margined to permit simultaneous reads from all ports. Simultaneous writes to the same address location are prohibited. All circuitry outside the core cell is replicated on a per port basis. Only the layout and tiling algorithms are unique for each RAM family.

Table 3: RAM Core Cell Area and Performance Comparison for 8k × 8 Block Sizes at Nominal Process, 27 °C and 5.0 V. Average power numbers are per port.

Type	Cell Area μm^2	tCHQV ns	tCHCH ns	Pave mW / MHz
4T DRAM	73.2	4.6	5.0	4.0
1 Port SRAM	131.3	4.4	4.6	4.2
2 Port SRAM	197.8	5.3	5.5	4.3
4 Port SRAM	397.8	5.4	5.6	4.5

DRAM IMPLEMENTATION

Although little has been published on modular DRAM design, embedding DRAM in place of SRAM can often reduce ASIC die size [8]. A modular DRAM based on a planar one transistor (1T) cell was developed for the Northern Telecom Electronics Ltd. 1.2 µm CMOS process. However, to meet alpha particle noise immunity specifications in BATMOS would require a planar 1T cell that is only marginally smaller than a 4T cell since no DRAM-style dense capacitor structures are available. The 4T cell was chosen since it offers considerable speed and power advantages and is compatible with the peripheral circuitry used in the rest of the RAM family.

The 4T DRAM cell uses the local interconnect layer in the BATMOS process to achieve an area which is 45% smaller than the 6T 1 port SRAM cell. The function of the 4T cell is identical to that of a 4T SRAM cell except for the refresh requirement. Cell refresh is achieved by the word line access operation. On the unselected columns, the active bit line precharge devices ensure the maximum cell 'one' level of V_{DD} - V_{TN} is achieved by clamping the bit line at V_{DD}. This 'one' level is also guaranteed on the accessed columns since the 'high' bit line potential is clamped to a minimum value of V_{DD} - V_{TN} by the write access devices.

RESULTS

Figure 4 shows a die photo of the core memory chip for a broadband time switch designed with the 2 port SRAM. The 96K (12K × 8) function was subdivided into four 3K × 8 blocks to minimize power. Both ports operate continuously at 100 Mhz dissipating 290 mW at nominal process, 5 V and 27 °C. Since the self-timed power down tracks process and temperature, the power dissipation at 100 MHz rises to only 340 mW at best case process, 5.5 V and 0 °C.

CONCLUSIONS

The design of a synchronous self-timed memory has been described. This architecture is ideally suited to modular memories embedded within synchronous systems due to its simple interface specification, excellent speed / power performance and ease of modeling. The basic port design is self-contained and extensible to any number of ports sharing access to a common core cell array. Implementations of 1, 2 and 4 port RAMs have been demonstrated with 5 ns nominal access and cycle times in a 0.8 µm BiCMOS process. Performance has been verified on a broadband time switch chip operating at a 100 Mhz clock rate. The same design has also been used to implement a modular DRAM based on a 4T cell which provides a 45% core cell density improvement over the 1 port SRAM with similar performance.

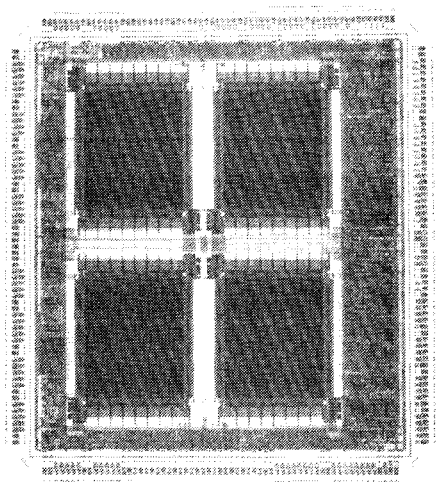

Figure 4: Die Photograph of a Time Switch ASIC Utilizing 96K of 2 Port SRAM

REFERENCES

[1] R. Hadaway *et al.*, "A Sub-Micron BiCMOS Technology for Telecommunications", *Proc. European Solid-State Device Research Conf.*, pp. 513-516, Montreux, 1991.

[2] B. Nadeau-Dostie, A. Silburt and V. Agarwal, "Serial Interfacing for Embedded Memory Testing", *IEEE Design and Test of Computers*, pp. 52-63, 1990.

[3] T. Akioka *et al.*, "A 6-ns 256-kb BiCMOS TTL SRAM", *IEEE J. Solid-State Circuits*, vol. 26, no. 3, pp. 439-443, March 1991.

[4] M. Takada *et al.*, "A 5-ns 1Mb ECL BiCMOS SRAM", *IEEE J. Solid-State Circuits*, vol. 25, no. 5, pp. 1057-1062, Oct. 1990.

[5] S. Shinagawa *et al.*, "A Multi-Speed Digital Cross-Connect Switching VLSI Using New Circuit Techniques in Dual Port RAMs" *Proc. IEEE Custom Integrated Circuits Conf.*, pp. 3.4.1-4, 1991.

[6] T. Dao and F. Sevejda, "A Dual-Port SRAM Compiler For 0.8-µm 100K BiCMOS Gate Arrays", *Proc. IEEE Custom Integrated Circuits Conf.*, pp. 22.4.1-3, 1991.

[7] H. Shinohara *et al.*, "A Flexible Multi-Port RAM compiler for Datapath", *Proc. IEEE Custom Integrated Circuits Conf.*, pp. 16.5.1-4, 1990.

[8] D. Lee and R. Katz, "Using Cache Mechanisms to Exploit Nonrefreshing DRAM's for On-Chip Memories", *IEEE J. Solid-State Circuits*, vol. 26, no. 4, pp. 657-661, April 1991.