

SRAMs strive to specialize

SRAM MANUFACTURERS ARE RESPONDING TO TOUGH BUSINESS CONDITIONS AND CHANGING CUSTOMER NEEDS WITH A RANGE OF DEVICES TARGETING SPECIFIC APPLICATIONS. THIS TREND MAY MEAN FEWER SUPPLY OPTIONS AND TOUGHER ARCHITECTURE DECISIONS, BUT THE END RESULT IS AN OPTIMIZED PART FOR YOUR UNIQUE REQUIREMENTS.

SO MANY CHIPS, SO FEW HOMES. These are rough times to be an SRAM vendor, and it's been this way since 1995, the peak of the last semiconductor boom. Companies are exhibiting mixed reactions to the current silicon slump: Some have quietly exited the SRAM business, refocusing their manufacturing lines on more lucrative logic devices. Others are staying the course with products targeting essentially the same markets as in the past, reducing costs when possible, and hoping that the flow of red ink ends soon.

A third group of vendors is directing its SRAM expertise away from PC and workstation caches and toward alternative applications. For example, the market for handheld communications and computing, or wireless systems, values low power consumption and is often willing to trade off speed to get it. Board space is another important consideration in wireless systems that is now driving numerous packaging innovations and that might in the

long run lead to an exodus from discrete chips and toward SRAM embedded on ASICs (see **sidebar** "The (SRAM) times, they are a-changin'").

Networking (data-communications) and telecommunications systems in many cases have exact opposite requirements. Gigabit Ethernet and asynchronous-transfer-mode (ATM) protocols accelerate SRAM performance demands, made especially challenging by the balanced proportion of

At a glance **64**
The (SRAM) times, they are a-changin' **66**
SRAM substitutes **68**
SRAM derivatives **72**
For more information **74**



SRAM

Dual-port
SRAM

NO-LATENCY
SRAM

LOW-POWER
SRAM

read and write operations and the high degree of random access. The 32- and 64-bit CPU and ASIC data interfaces require wide memory buses, zero-downtime operation leads to the need for parity support, and more numerous and richer data types drive higher densities. Here too, bit growth and unrelenting cost pressures may in the long term lead to SRAM's demise, this time at the hands of low-latency DRAM (see sidebar "SRAM substitutes").

STANDARD PRODUCTS

In the beginning (and still today), there were asynchronous SRAMs, with generic interfaces containing address and data buses, read and write strobes, and chip selects. Asynchronous-SRAM speeds have decreased over time from 10s or hundreds of nanoseconds to today's less-than-10-nsec offerings. As data buses get wider than 8 bits, vendors have added byte-enable inputs to allow for selective data alteration.

The first significant SRAM-interface innovation occurred with the unveiling of Intel's (www.intel.com) i486 and Motorola's 68040 μ Ps, which represented the first high-volume CPUs with a burst local bus.

AT A GLANCE

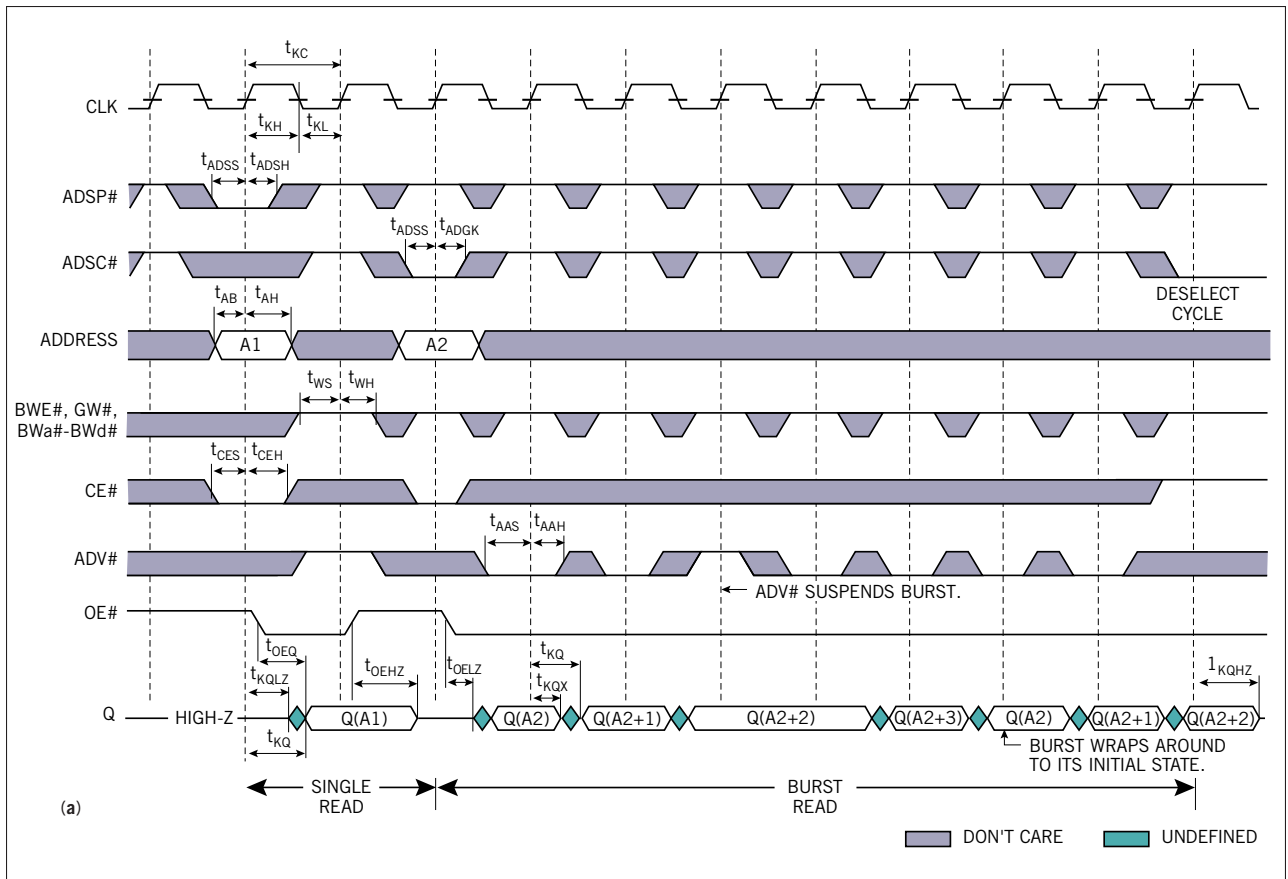
- ▶ Uncertain economics and evolving applications are behind the changes under way in SRAMs.
- ▶ Pipelined and late-write SRAMs remain dominant in computing; no-latency follow-ons target networking.
- ▶ Low power consumption is not the only motivating factor in the move to six-transistor cells.
- ▶ SRAM derivatives solve specific requirements, and embedded arrays and DRAM-based memories may replace some of today's SRAM business.

SRAM manufacturers responded by modifying their devices' interfaces for synchronous operation and, in some cases, integrating a multibit burst-address counter that supports both interleaved and linear burst orders via control input polarity.

Today's synchronous-SRAM data buses are either flow-through or pipelined (Fig-

ure 1). Both alternatives latch addresses and control signal states on the rising edge of the input clock. Both SRAM versions also have similar write-access profiles. However, the two alternatives differ in their read-data-output latencies and maximum operating speeds. Flow-through synchronous SRAMs output their read data during the clock period immediately after the rising edge, which latches in address and control signals. The system typically latches this output data on the next rising clock edge.

Pipelined synchronous SRAMs add a bank of registers after the device's sense-amp stage and before its output buffers. These registers create an additional one-clock latency before the device outputs its first valid data. However, by spreading the read delay across multiple clocks and by allowing sense and output portions of consecutive reads to occur in parallel, pipelined synchronous SRAMs can run at higher clock frequencies than flow-through alternatives. At high clock rates and with long burst cycles, which amortize the additional one-clock initial latency, pipelined parts deliver not only higher burst but also



higher sustained bandwidth. On the other hand, flow-through interfaces enable the fastest transition from read to write on a given I/O port.

Although intended for desktop and notebook PCs, synchronous SRAM has traditionally delivered adequate performance for other applications, especially considering the low memory costs and abundant number of suppliers driven by the PC's high volumes. However, keep in mind the usage environment with 486-class (and even Pentium-class) systems. The DOS and DOS-based kernels comfortably fit within the L1-plus-L2 cache in many cases (min-

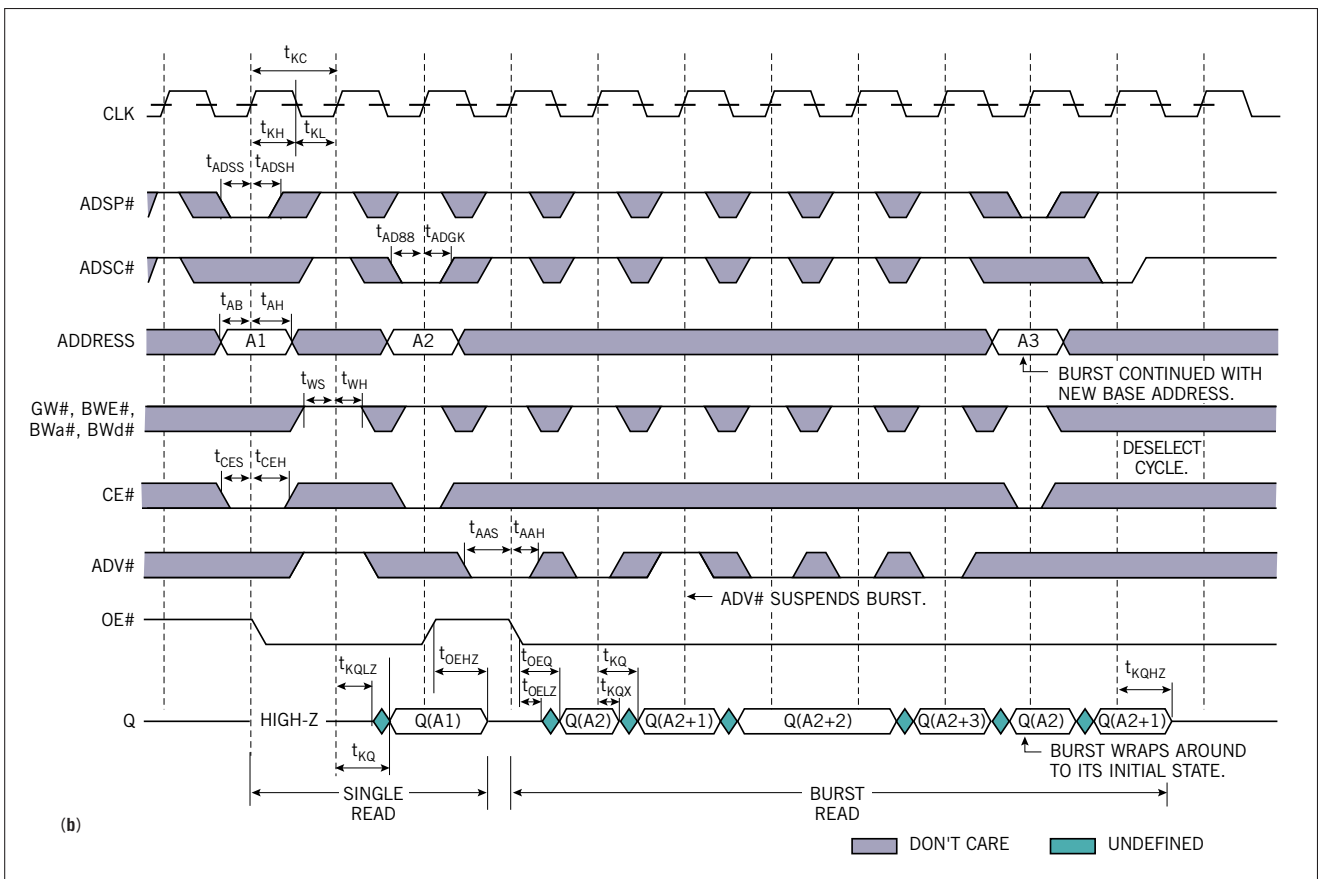
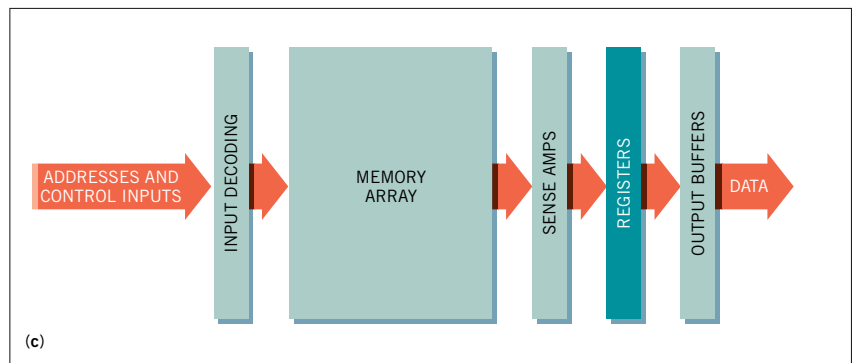
imizing the percentage of writes to SRAM), applications frequently ran in a single-task fashion, data types were relatively small and simple, and bursts were usually four or eight accesses long (for L1-cache line fills).

Workstations and servers presented a more complicated scenario. They use comparatively elaborate operating systems and contend with multiple multitasking applications and numerous client transactions, both manipulating more complex data sets. Also, the data-bus widths from μP to external cache often weren't $\times 32$ or $\times 64$, as in standard PCs, as were $\times 128$ (plus parity). As a result, burst lengths for on-CPU

L1-cache fills were shorter, increasing the negative performance impact of any access latencies. The two clocks of turnaround time between a write and a subsequent read proved to be an unacceptable performance limiter. In response, a group of SRAM users and manufacturers in 1992 formed the Motorola SRAM Users Group (MSUG) and developed the late-write synchronous SRAM architecture (Figure 2, pg 70).

Late-write synchronous SRAMs also come in flow-through and pipelined versions and reduce but do not eliminate the write-to-read clock latency of standard synchronous SRAMs. They specify an op-

Figure 1 Flow-through synchronous SRAMs in (a) (bottom of pg 64) enable fast bus turnaround between reads and writes, whereas pipelined alternatives (b) exhibit an additional one-clock read-access latency because of their on-chip register bank (c) but frequently run at higher clock frequencies (a and b courtesy G-Link Technology).



THE (SRAM) TIMES, THEY ARE A-CHANGIN'

The mid-'90s were good years to be an SRAM vendor (or in general, *any* type of semiconductor manufacturer). Intel's (www.intel.com) Pentium microprocessor and Microsoft's

(www.microsoft.com) Windows 95 operating system combined to fuel demand for both new and replacement PCs, resulting in a scarce supply of L2-cache chips (and, correspondingly, high prices). A number of suppliers refocused their fabrication facilities to produce SRAM, and brand-new fabs opened worldwide. The resulting supply glut resulted in the current downward price spiral that began in late 1995.

Supply is only half of the supply/demand relationship that defines prices in an open market. Unfortunately for the SRAM vendors, PC demand for SRAM as L2 cache seems to be rapidly disappearing. When Intel moved its CPUs from Socket 7 (Pentium) to Slot 1 (Pentium II) packaging, it began buying the SRAM chips itself for integration onto the CPU modules. Previously, the SRAMs came on separate memory boards, which PC manufacturers purchased and installed, along with the CPU and other components and add-in cards, on the motherboard.

Only a few SRAM companies supply Pentium II L2 cache. Intel does not identify them, and nondisclosure agreements prohibit the vendors from identifying themselves, but an often-repeated rumor indicates that the companies are Mitsubishi, Motorola, NEC, Samsung, and Toshiba. Many other vendors

became concerned with the large number of suppliers pursuing one—albeit large—customer. The downward spiraling prices that resulted caused many of these vendors to exit the PC L2-cache business. The continuing presence in the Socket 7 CPU market of AMD (www.amd.com), Cyrix (www.cyrix.com), and Inte-

ability wouldn't impact the production ramp of the Pentium Pro and Xeon CPUs. The company also claims that it is unopposed to outsourcing cache in the future if other suppliers prove they can achieve the required performance goals in sufficiently high volume and with low enough cost to meet Intel's needs. Intel doesn't comment on SRAM's effectiveness as a fab filler during times of lower than forecasted CPU demand.

The long-term trend, though, is clearly to integrate L2 cache onto the CPU die. This transition is the next logi-

cal step that began with the L1 cache and floating-point unit in 486 CPUs. Intel's second-generation Celeron μ P (Figure A) is the first example of this trend, beginning at the low end of the performance range, which requires small and, therefore, easily integrated caches, and migrating upward over time.

Ironically, although the integrated L2 cache may be smaller than it could be if designers implemented it externally, its higher performance resulting from reduced access delays, wider on-chip buses, and more exotic configurations may give it performance comparable with the off-chip alternative. High-bandwidth Direct Rambus DRAMs, due to appear next year, will further reduce the need for large, fast L2 external cache. Workstations and servers based on non-x86 CPUs, such as Alpha, MIPS, PA-RISC, and SPARC, seem content for now with off-chip L2 cache, but the migration onto the CPU die is probably only a matter of time.

The other area of today's SRAM business most vulnerable to integration is the low-power segment. Embedding SRAM, along with analog circuitry and digital logic, onto an ASIC provides several advantages for portable equipment. First, by reducing chip count, you minimize board space (an obvious motivation to anyone who's seen the latest versions of digital cellular phones). Also, by not having to drive an external interface between logic and memory, you can noticeably reduce total power consumption, something that NeoMagic (www.neomagic.com) figured out a few years back with notebook-PC graphics controllers (Reference A).

Several factors may slow this migration to embedded SRAM, however. The extra design and debugging effort that integrating the memory array requires is at odds with the continual pressure to reduce system-design schedules. With SRAM components so cheap nowadays, an embedded memory in many cases will result in a more expensive total solution. Ultimately, factors other than the board area, such as the keypad, LCD, speaker, and microphone, constrain the ability to shrink the size of a cellular phone, pager, or personal digital assistant. Finally, each feature you add—phone, pager, address book, Global Positioning System, games, or high-resolution display—all require more program and data memory and reduce the probability that embedded SRAM will be practical.

Reference

A. Dipert, Brian, "Embedded memory: the all-purpose core," *EDN*, March 13, 1998, pg 34.

Figure A



Compared with its Pentium II counterpart (left), the second-generation Celeron processor (right) integrates 128 kbytes of L2 cache alongside the 32-kbyte L1 and exemplifies an embedded-memory trend that may extend beyond the PC in the future (courtesy Intel).

grated Device Technology ensures other PC-cache opportunities, but for how long is anyone's guess.

Intel has been quietly manufacturing its own SRAM chips for several years, first for its Pentium Pro and now for its Xeon microprocessors. The company uses the double-data-rate interface to enable the cache data-transfer rate to match the core CPU operating frequency; with its standard Pentium II processors, Intel runs the external L2 cache at half the core frequency. Intel claims it chose to design and manufacture its own cache chips to ensure that their avail-

tional differentially driven clock, a requirement for highest speed designs but a questionable cost addition at lower speeds (and a complication to system interfacing if other components use a standard clock). With operating frequencies greater than 300 MHz, late-write SRAMs from companies such as Hitachi, IBM, Motorola, and Sony remain common in high-end computers. Some networking and telecommunications designs are also adopting late-write SRAMs, but an even lower latency alternative is becoming popular in these applications.

THE NEED FOR SPEED

Except in highest end configurations, datacomm and telecomm systems often rely on DRAM in conjunction with on-CPU L1 cache (if available) for code execution. SRAM therefore finds uses as a data buffer between input and output ports and as a high-speed look-up table containing addresses and other information to route data streams from source to destination in the communications network. Reducing or eliminating write-to-read delays is crucial to maximizing system performance.

Integrated Device Technology (IDT) two years ago introduced its first zero-bus-turnaround (ZBT) synchronous SRAM in a 1-Mbit density. No-latency SRAMs such as the ZBT architecture deliver 100% databus efficiency, assuming that the system has previously filled the address pipeline. Like late-write SRAMs, no-latency devices latch written data on a different clock rising edge from the one that latches addresses and control signals. On the other hand, no-latency SRAMs use a standard clock with all input and output signal timings specified as setups and holds relative to the rising edge. Just as with standard and late-write synchronous SRAMs, no-latency devices come in both flow-through and pipelined versions (Table 1).

Although IDT's first-generation device wasn't a significant market success, the company has high hopes for its 4-Mbit ZBT memories. The company built these devices on a more advanced lithography, and the memories' larger arrays reduce the percentage cost of on-chip ZBT logic. Among other functions, this logic handles the scenario you encounter when you attempt to immediately read from a location you've just written to before the chip has sufficient time to modify the appropriate

TABLE 1—COMPARISON OF PBSRAM, LATE-WRITE SRAM, AND NO-LATENCY ACCESS PROFILES

Memory	No. of clocks ¹	Clock profile ¹
PBSRAM (pipelined)	16	3-1-3-1-3-1-3-1
Late write (pipelined)	13	3-1-2-1-2-1-2-1
NoBL/ZBT (pipelined)	10	3-1-1-1-1-1-1-1
NoBL/ZBT (flow through)	9	2-1-1-1-1-1-1-1

¹Eight-access sequence consists of read-write-read-write-read-write-read-write.

array cells. An on-chip latch-and-multiplexer circuit routes the just-written information back to the data bus.

IDT has entered into an alternative-sourcing agreement with Micron Technology and Motorola. The three suppliers' parts may differ from each other in minor respects because three design teams created them and the companies fabricate them on unique semiconductor processes. However, the companies at least discuss future product ideas and agree on common packages, pinouts, functions, and data-sheet specifications for SRAM densities currently as large as 16 Mbits. Other suppliers have jumped on the no-latency bandwagon, but you might not be able to tell this fact from a quick glance at their Web sites or data books.

Alliance Semiconductor's No-Wait, Cypress Semiconductor's NoBL (no bus latency), G-Link Technology's Dual Late

Write, Samsung's NtRAM (No Turnaround RAM), Sony's Late Late Write, Toshiba's Double Late Write, Winbond's Zero Wait State, and others all implement the same no-write-to-read latency concept. However, all these names allow the vendors to avoid ZBT trademark infringement and to attempt to differentiate their "brands" in the crowded, hypercompetitive SRAM market. Although the vendors all claim compatibility between their memories and those of other manufacturers, it's doubtful that any share advance specifications with each other before product introduction. You should carefully examine each vendor's data sheets and, if necessary, adjust your design if you want the flexibility of multiple suppliers. The largest no-latency SRAMs currently come in 4-Mbit densities, although Samsung, Sony, and Toshiba all plan to offer samples of 8-Mbit versions by year's end.

At high clock rates, bus contention becomes a concern when you attempt to switch between read and write operations without intermediate "dead" clock cycles. Will the system quit driving the data bus before the memory's output buffers turn on (write to read), and will the SRAM's outputs float before the ASIC or CPU ac-

(continued on pg 76)

SRAM SUBSTITUTES

Your memory-density needs are increasing, but the downward cost pressure is just as unrelenting. You've so far used SRAM, but you're unsure that it will let you hit your design targets this time. What do you do? Consider DRAM.

A variety of high-performance DRAMs are in production, and more are in development. They include commodity SDRAM, currently supporting the PC-100 specification; synchronous-graphics RAM (SGRAM) as fast as 200 MHz; upcoming double-data-rate versions of both SDRAM and SGRAM; and the Rambus (www.rambus.com)-defined standard for concurrent and direct RDRAM. Vendor-proprietary DRAM enhancements,

offering low latency, high bandwidth, or both, include Enhanced Memory's (www.edram.com) enhanced DRAM and enhanced SDRAM, Fujitsu's (www.fujitsumicro.com) fast-cycle RAM, Mitsubishi's cache DRAM, MoSys' MDRAM, and NEC's virtual-channel DRAM (refer-ences A and B).

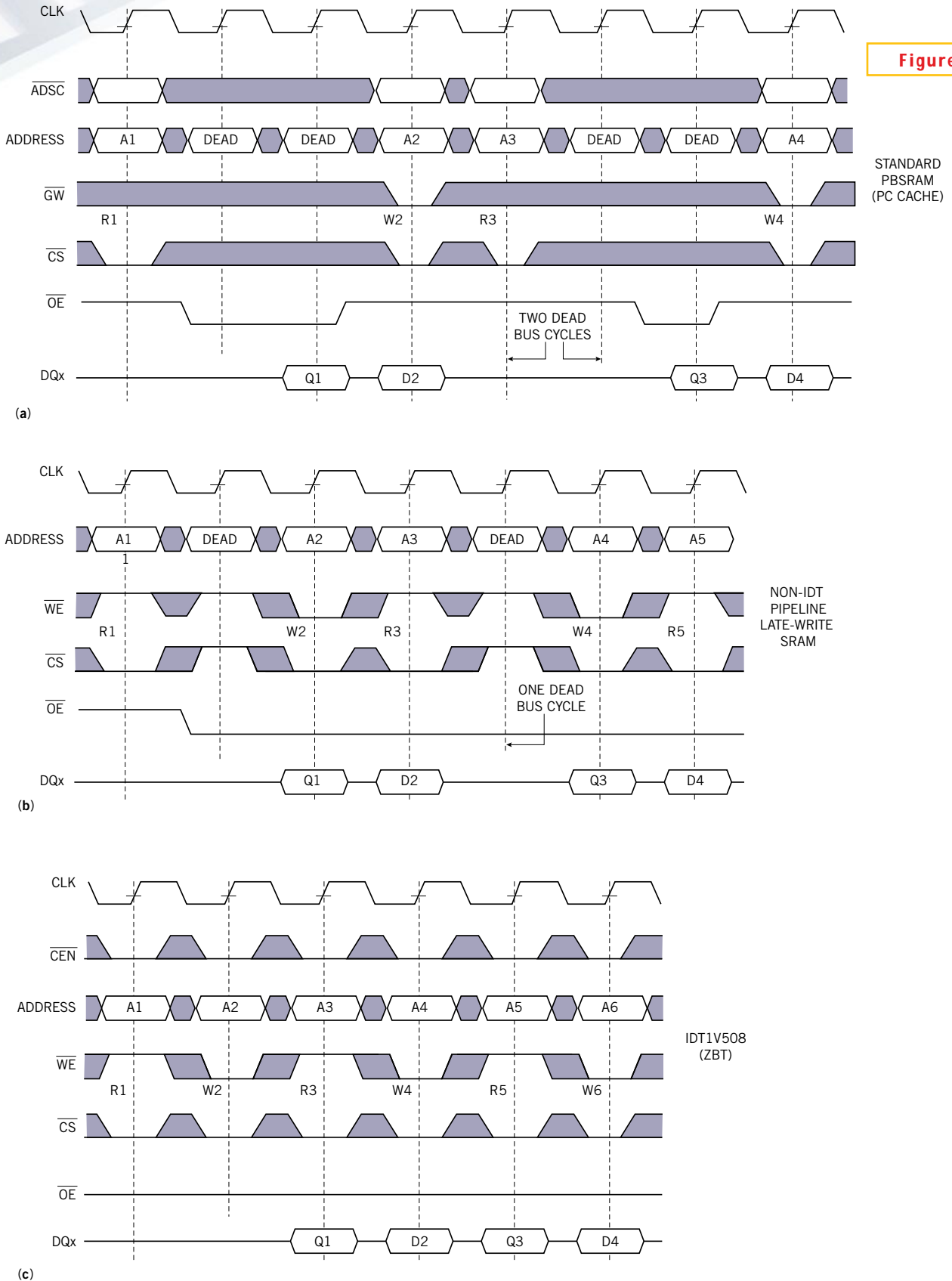
MoSys has taken DRAM replacement one step further. The company's MCache architecture wraps an SRAM-like system interface around the MDRAM array core. The recently announced one-transistor SRAM removes the need for external refresh control, which the first-generation MCache required. And, according to the vendor, one-

transistor SRAM fully emulates synchronous burst SRAM but doesn't offer full random access-to-access, zero-wait-state performance. Mosis is offering samples of 2- and 4-Mbit chips that run as fast as 133 MHz and plans higher speed versions and no-latency derivatives (see the Leading Edge article, "SRAMs take a new approach" on pg 18 in this issue).

REFERENCES

- A. Dipert, Brian, "DRAM gets to the core of the problem," *EDN*, Jan 1, 1998, pg 14.
- B. Dipert, Brian, "DRAM redesign: not just plastic surgery," *EDN*, July 2, 1998, pg 20.

Figure 2



Standard (a), late-write (b), and no-latency (c) synchronous SRAMs make incremental improvements in sustained data-bus bandwidth when the system switches between write and read operations (courtesy Integrated Device Technology).

SRAM DERIVATIVES

IN NO OTHER SEMICONDUCTOR product segment is the conflict between commodity and specialized devices more noticeable than in memories.

This situation is the result of both the large volume of units shipped each year and the total market size. Vendors can try to make money by selling industry-standard devices, which attracts both many customers and many competitors. Such vendors hope to make at least a small profit margin per chip, balanced out by high unit volumes.

Another way for vendors to make money is to sell more specialized memories. Few customers will buy them because they have only one or a few supply sources. But, if vendors do a good job of defining the devices, their customers will pay a premium. In theory—and sometimes in reality—the vendors sell fewer chips but make more money on each one. Supply/demand imbalances also somewhat insulate these vendors' prices from fluctuations in the commodity-memory market. Some companies try to sell both commodity and differentiated memories, but few succeed because the two options require fundamentally different business models.

Multiport SRAMs provide an example of a differentiated memory. Their array cells build on a standard SRAM heritage with additional transistors per bit to enable access from multiple data-bus sources (Figure A). Multiport SRAMs conceptually serve similar application requirements as do FIFO buffers—as bridges between buses operating at different speeds, with different burst lengths, or with different widths. However, multiport

SRAMs offer more functions than their FIFO counterparts.

Unlike unidirectional FIFO buffers, multiport SRAMs let you read *and* write both ports. They also enable full random access of any location in the memory array, whereas both FIFO and last-in-first-out (LIFO) access protocols imply sequential “streamed” data patterns. The multiport SRAMs also find use when two hosts, such as two DSPs or a local processor and a global interface bus, need to simultaneously

read or write different memory locations.

Cypress Semiconductor and Integrated Device Technology (IDT) are the main vendors supplying multiport SRAMs—unfortunately, with pinout- and function-incompatible devices. IDT also sells bank-switchable dual-port SRAMs and serial-access RAMs, and Quality Semiconductor sells shared-port SRAMs, all of which offer lower flexibility but claim lower cost. In comparing multiport memories for use in your design, you should consider the following questions:

- Does your design require asynchronous ports, synchronous ports, or both? If synchronous, do flow-

through, pipelined, or both modes require support?

- Can one device offer sufficient interface flexibility for your needs, and are you paying extra for flexibility you won't use?
- How many ports do you need?
- Is parity support valuable?
- Is package and pinout compatibility between various densities and operating-voltage ranges necessary?
- How important are additional logic resources, such as on-chip arbitration circuitry or burst address counters?
- Can you easily expand the memory array's depth or width without a large amount of separate glue logic?

Content-addressable memory (CAM), another differentiat-

ed device based on SRAM technology, conceptually works in a backward fashion compared with standard memories. Instead of your supplying CAMs an address and their returning the data at that address, you give them a seek-function data sequence, and they return both the data's address and additional associated data. For writes, CAMs use no addresses. Instead, a learn function writes information to the CAM, and it uses whatever array locations are available.

In the CAM equivalent of an SRAM read, you first configure one or several integrated mask registers and then supply the *comparand*, a data-bit sequence. Based on the comparand and mask-register values, the CAM *simultaneously* compares each array location's bits with the comparand. It can perform this task because each CAM bit is a 10-transistor or an eight-trans-

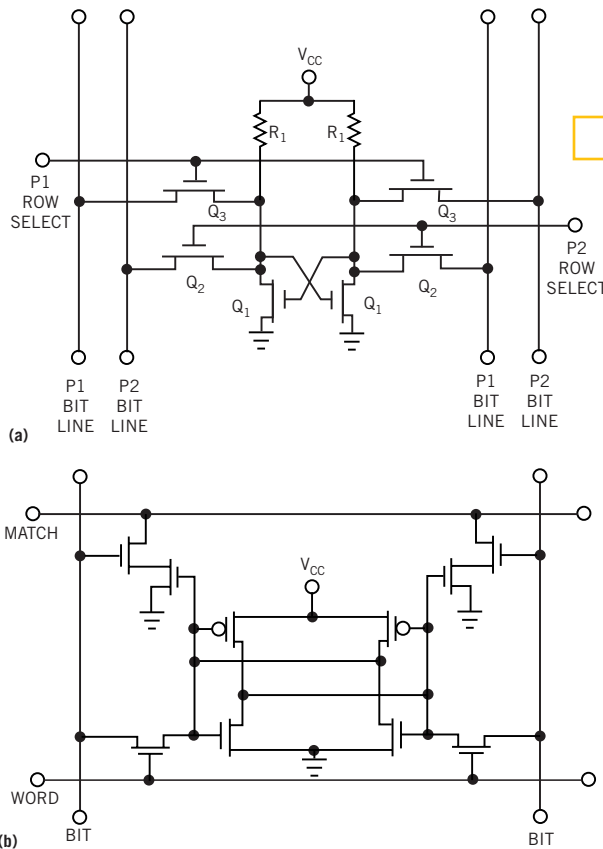
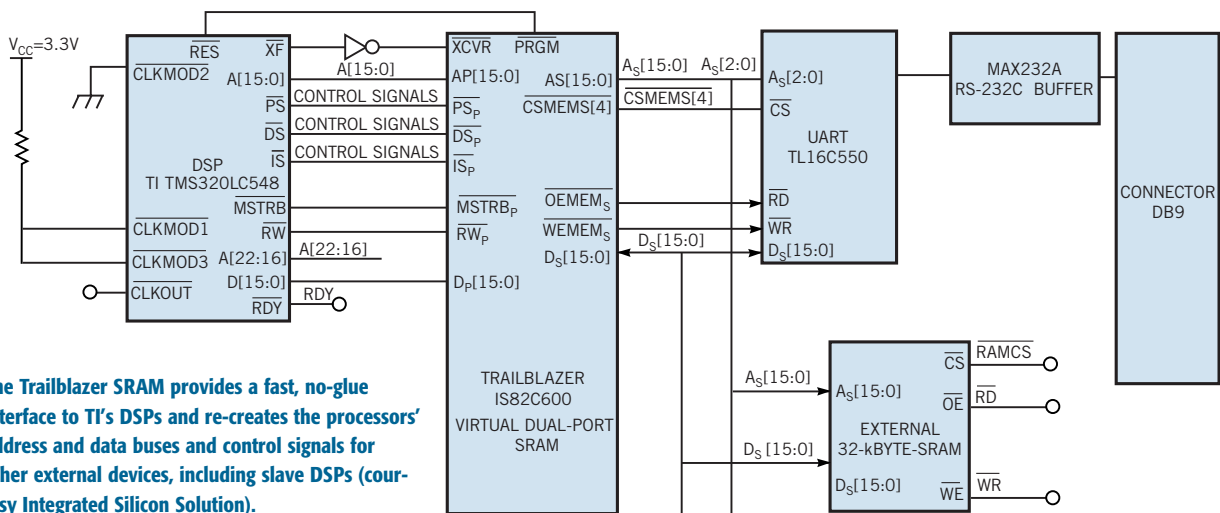


Figure A

A close analysis of multiport (a) and CAM (b) cells reveals their SRAM heritage (a, courtesy Integrated Device Technology; b, courtesy Music Semiconductor).



The Trailblazer SRAM provides a fast, no-glue interface to TI's DSPs and re-creates the processors' address and data buses and control signals for other external devices, including slave DSPs (courtesy Integrated Silicon Solution).

sistor, two-resistor cell with the four additional transistors beyond an SRAM implementing an XNOR function (Figure B). When the CAM finds a match, it outputs the match-detected status (a wired-or of all XNOR outputs), the match's address and its "associated data," and the remainder of the bit sequence at that location. Because the mask registers let you set X or don't-care values, they potentially create a match with multiple array locations. In this case, the CAM reports a multiple-match status and commonly first supplies the associated data at the lowest address. You can then read out the remainder of the matches one at a time.

Just as with multiport SRAMs, a variety of incompatible CAM devices is available from companies such as Kawasaki LSI (www.klsi.com), Lara Technology (www.laratech.com), Music Semiconductor (www.music.com), NetLogic Systems (www.netlogic.com), and Quality Semiconductor. Evaluation criteria you should consider include:

- memory density and internal data-bus width,
- ability to easily extend depth and width by chain-

- ing together multiple memories,
- seek speed and whether this speed includes "learning" the new data if the memory finds no match,
- number of mask registers (some devices even offer one mask register per CAM location),
- presence of separate comparand, result, and command ports to maximize total data bandwidth, and
- bus width of each interface port, especially as it compares with internal bus width.

Motorola offers CAM-like and multiportlike memories based on standard SRAM cells, trading off reduced functionality and performance for claimed lower cost. The company's Binary CAMs use a hashing technique that stores data in increasing values, beginning with the lowest order SRAM address. This approach means that each data update requires a reorder of an average of half the SRAM contents. When searching for a data match, the internal logic begins halfway through the SRAM array and sequentially zeroes in on the match, using repeated greater than and less than comparison results. Binary CAMs integrate the external

logic that yet another CAM alternative (employing a standard SRAM and separate microcontroller, ASIC, or programmable-logic device) uses. Hashing provides no consistent, deterministic search delay. (Some searches complete faster than others.) However, you can still calculate the maximum delay latency, which may be sufficient for your application. Motorola's NetRAMs externally appear to work just like standard synchronous dual-port SRAMs; however, NetRAMs internally read and write one port on one edge of the input clock and the other port on the opposite edge. You should consider NetRAMs if both of your system buses operate from the same clock. However, if a port needs to operate asynchronously and if the ports run at different frequencies or phases, you need a true multiport SRAM. Multiport SRAMs and CAMs

are complex and interesting architectures. Look for a feature article on them next year in EDN. Some derivative memories are less radical in their departure from standard products; Integrated Silicon Solution (ISSI) and Motorola both supply 3-Mbit, $\times 24$ interface SRAMs for Motorola's DSPs. Others offer unique features. For example, ISSI's \$10 (10,000), 64kbit $\times 16$ Trailblazer offers a no-glue, zero-wait-state, point-to-point connection to TI's TMS320LC5x and TMS320LC54x DSPs at speeds as high as 133 MHz and re-creates the TI DSP's address, data, and control buses for interfacing to other system chips (Figure A). This interfacing includes the ability to act as a bridge between multiple processors by broadcasting all cycles or all-except-SRAM cycles from the primary to the secondary DSP port.

Figure B

FOR MORE INFORMATION...

For more information on products such as those discussed in this article, circle the appropriate numbers on the Information Retrieval Service card or use EDN's InfoAccess service. When you contact any of the following manufacturers directly, please let them know you read about their products in EDN.

Alliance Semiconductor Corp

San Jose, CA
1-408-383-4900
fax 1-408-383-4999
www.alsc.com
Circle No. 424

Aptos Semiconductor Corp

San Jose, CA
1-408-474-3000
fax 1-408-474-0445
www.aptos.com
Circle No. 425

Cypress Semiconductor

San Jose, CA
1-408-943-2600
fax 1-408-943-2741
www.cypress.com
Circle No. 426

Enable Semiconductor Inc

Milpitas, CA
1-408-955-9888
fax 1-408-955-9887
www.enablesemi.com
Circle No. 427

Etron Technology Inc

Santa Clara, CA
1-408-987-2255
fax 1-408-987-2250
www.etrn.com
Circle No. 428

G-Link Technology

Santa Clara, CA
1-408-492-9068
fax 1-408-492-9067
www.glinktech.com
Circle No. 429

Galvantech Inc

Santa Clara, CA
1-408-566-0688
fax 1-408-566-0699
www.galvantech.com
Circle No. 430

Hitachi Semiconductor America Inc

Brisbane, CA
1-650-589-8300
fax 1-650-583-4207
www.hitachi.com/semiconductor
Circle No. 431

Hyundai Electronics America

San Jose, CA
1-408-232-8000
fax 1-408-232-8125
www.he.com
Circle No. 432

IBM Microelectronics Corp

Armonk, NY
1-914-765-1900
fax 1-914-892-5334
www.chips.ibm.com
Circle No. 433

Integrated Device Technology Inc

Santa Clara, CA
1-408-727-6116
fax 1-408-492-8674
www.idt.com
Circle No. 434

Integrated Silicon Solution Inc

Santa Clara, CA
1-408-588-0800
fax 1-408-588-0806
www.issiusa.com
Circle No. 435

LG Semicon

San Jose, CA
1-408-432-5000
fax 1-408-432-6067
www.lg.co.kr
Circle No. 436

Micron Technologies

Boise, ID
1-208-368-3900
fax 1-208-368-4431
www.micron.com
Circle No. 437

Mitsubishi Electronics America Inc

Sunnyvale, CA
1-408-730-5900
fax 1-408-732-9382
www.mitsubishichips.com
Circle No. 438

Mosaic Semiconductor Inc

San Diego, CA
1-619-271-4565
fax 1-619-271-6058
www.syntaq.com
Circle No. 439

Mosel Vitelic Inc

San Jose, CA
1-408-433-6000
fax 1-408-433-0952
www.moselvitelic.com
Circle No. 440

MoSys Inc

Sunnyvale, CA
1-408-731-1800
fax 1-408-731-1893
www.mosys.com
Circle No. 441

Motorola Inc

Austin, TX
1-800-521-6274
www.motorola.com/sps
Circle No. 442

NEC Electronics Inc

Santa Clara, CA
1-408-986-1020
fax 1-408-588-6374
www.nec.com
Circle No. 443

Paradigm Technology Inc

Milpitas, CA
1-408-954-0500
fax 1-408-954-8913
www.prdm.com
Circle No. 444

Quality Semiconductor Inc

Santa Clara, CA
1-408-450-8000
fax 1-408-954-8913
www.qualitysemi.com
Circle No. 445

Samsung Semiconductor

San Jose, CA
1-408-954-7000
fax 1-408-954-7870
www.samsung.com
Circle No. 446

Sony Semiconductor Corp of America

San Jose, CA
1-408-955-6572
fax 1-408-955-5176
www.sony.com/semi
Circle No. 447

Synergy Semiconductor

San Jose, CA
1-408-980-9191
fax 1-408-914-7878
www.synergyssemi.com
Circle No. 448

Toshiba America Electronic Components Inc

Irvine, CA
1-714-455-2000
fax 1-714-859-3963
www.toshiba.com/taec
Circle No. 449

Winbond Electronics Corp

Hsinchu, Taiwan
886-3-5770066
fax 886-3-5792668
www.winbond.com.tw
Circle No. 450

SUPER CIRCLE NUMBER

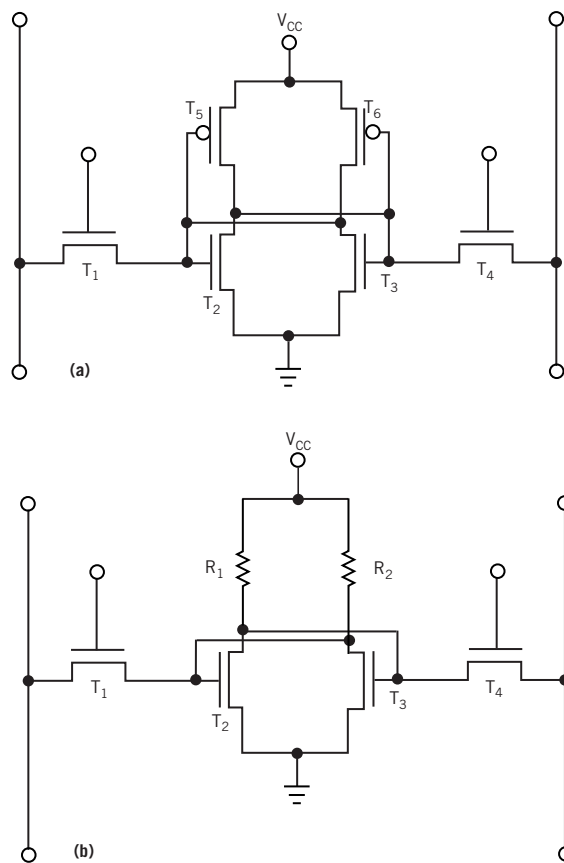
For more information on the products available from all of the vendors listed in this box, circle one number on the reader service card. Circle No. 451

tivates (read to write)? Casual analysis of data sheets might produce an unnecessarily pessimistic outlook, because the minimum output-buffer turn-on time and maximum output-buffer turn-off time typically occur at opposite ends of the operating-voltage and -temperature range (Reference 1).

SRAM manufacturers claim that no-latency synchronous-SRAM operation as fast as 150 MHz or so should be no problem, and they are working on faster turn-off output buffers (and hoping the ASIC manufacturers are doing the same) to enable boosting the clock rates even higher. Alternatively, a few companies are considering dual-data-bus parts—one for input, the other for output—although the added pin count would make for an expensive chip, even in BGA packaging (see sidebar “SRAM derivatives”).

Yet another, even higher performance RAM-interface option looms on the horizon with possible application to standard, late-write, and no-latency architectures. Double-data-rate (DDR) SRAMs transfer data on both edges of the input clock, meaning that they have double the instantaneous bandwidth of single-data-rate devices. The DDR devices’ sustained bandwidth, however, is always less than two times that of single-data-rate devices, because address and control signals (which are most susceptible to device-to-device skew effects) transfer on only one clock edge. MSUG has published a first-generation specification for DDR SRAM and is finalizing a second-generation document with modified read-data echo-clock functions.

A number of SRAM vendors report that they’re in either the product-definition or the early design stages of DDR SRAM. They agree that successful completion of in-progress industry standardization efforts and adoption by multiple compatible supplier sources are requirements to ensure that the architecture achieves high-volume, long-term success. A few companies have even developed early DDR-



Six-transistor SRAMs (a) have larger cells yet provide lower power consumption, lower operating voltages, and better reliability than their four-transistor, two-resistor counterparts (b). At advanced lithographies, the cell-size differences become negligible (courtesy Cypress Semiconductor).

SRAM prototype versions, which they are offering for sampling to potential customers for evaluation and feedback. DDR SRAMs, like DDR SDRAMs, threaten to be a significant system-design challenge over their single-data-rate predecessors but offer the most realistic means of pushing SRAM-component performance significantly beyond 300 MHz (Reference 2).

What about on-chip copper interconnection as a performance booster? This topic, along with silicon-on-insulator (SOI) technology, has recently become popular in the electronics press and reflects a growing amount of research and development by a number of companies (Reference 3). Copper, a lower impedance alternative to the traditionally used aluminum, provides the potential to reduce on-chip routing delays when all other factors are equal. Several companies are running prototype copper-based SRAM lines in their development fabrication facilities because SRAM shares a high degree of

process compatibility with logic and because the highly repeatable SRAM array simplifies the detection and elimination of randomly located fabrication-defect mechanisms. However, copper’s applicability to improving SRAM speeds in high-volume production is unclear.

At a given routing-channel width, copper offers a lower impedance (and therefore faster) pathway to signals. Instead of producing faster chips, however, some companies might reduce the routing widths or—even more drastic—the number of metal layers to lower die size and fabrication cost and still deliver aluminumlike timings. SRAM performance also depends on factors other than routing delays, including input- and output-buffer latencies and sense-amplifier speed. Also, the copper-based process presents greater production cost, scheduling, and volume risks for mainstream SRAMs than does the decades-old aluminum-based counterpart. Motorola is one vendor whose designers apparently think copper-based

SRAMs make sense. The company is now making an 8-Mbit, late-write SRAM operating as fast as 333 MHz and with $\times 18$ and $\times 36$ data-bus options.

WIDER IS BETTER

The number of nodes in networks and the amount of data transferred between these nodes is increasing, and the tolerance for long and unpredictable response latencies is decreasing with evolutions in data type. As a result, SRAM-density requirements are also growing at a healthy rate. Wider CPU and ASIC “pipes” also require wider SRAM data buses, and high clock rates increase the probability of noise-induced data errors. With high reliability a key requirement for networking systems, parity support is becoming commonplace. Many SRAM vendors are now offering 32- and 36-bit wide data buses.

To keep power consumption at reasonable levels, improve output-buffer switching speeds, and minimize noise effects, bus-

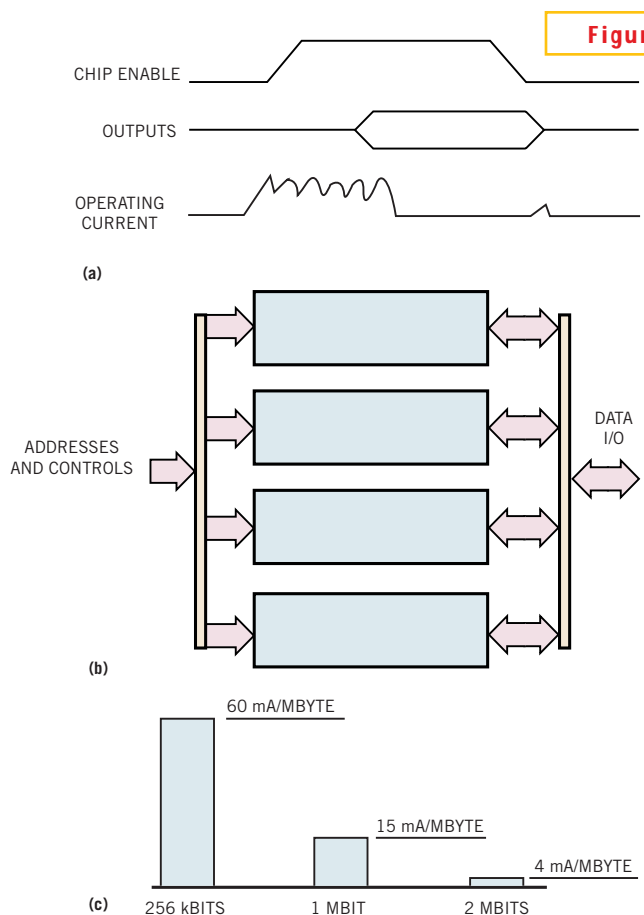
interface protocols are evolving away from the traditional 5V TTL and CMOS levels. Low-voltage TTL (LVTTTL) and low-voltage CMOS (LVCMOS) based on reduced-swing 0-to-3V transitions are now common, and high-performance parts use the 2.5V high-speed stub-terminated logic (HSTL) protocol. Even more esoteric interfaces exist for specialized applications, such as the Gunning transistor logic (GTL) I/O buffers that the backside bus of Intel's Pentium Pro uses. DDR SRAMs will probably use the stub-series-terminated logic (SSTL) interface, like their DDR SDRAM brethren.

In portable-equipment applications, such as cellular phones, Global Positioning System terminals, personal digital assistants, and pagers, low power consumption is often the most important selection criterion. This factor is one of the key motivations behind a widespread conversion from four-transistor, two-resistor (4T2R) to six-transistor (6T) cell structures (Figure 3). Historically, 4T2R-based SRAMs have been most common. Their smaller cells reduced

memory costs because the resistors stack above the transistors and therefore don't take up incremental die area. However, the passive pullup/pulldown resistors generate constant current flow in each cell, even with the entire memory deselected, an effect that becomes more pronounced as array densities increase.

The 6T SRAMs replace the stacked resistors with CMOS substrate-based transistors, which draw significant current only when they're switching (for example, during write operations and only for those data bits being altered). Also, 4T2R SRAMs require the deposition of an additional polysilicon layer as part of fabrication, a step that the migration to 6T structures would eliminate. Finally, the 4T2R cell's resistors don't easily scale downward with shrinks in lithography, which means that at some point, a 6T cell will be *smaller* than its 4T2R equivalent.

An alternative, less common SRAM-cell approach creates a hybrid 6T cell structure using amorphous silicon-based thin-film transistors (TFTs) above



SRAM manufacturers commonly power down portions of the memory's internal circuitry (a) and subdivide the array (b) to reduce active and standby power consumption (c) (courtesy Enable Semiconductor).

the substrate. The TFT technique theoretically combines the best of both 4T2R (size) and pure 6T (power, reliability) structures. However, TFTs have thus far proved to be a challenge to manufacture in high volumes and with consistent results, especially at high densities. Unless TFT predictability improves, SRAMs based on hybrid 6T cells will be unable to migrate to low-voltage and extended-temperature operation.

Cypress Semiconductor, with its 0.5- μ m RAM3 and follow-on processes, has executed an aggressive conversion to the 6T cell across all of its new SRAM-based products, whether or not they target low-power applications. All systems value high reliability, and the 6T cells offer greater cell stability in resisting alpha-particle effects. Cypress has also shrunk the 6T cell at a given lithography by using a self-aligned contact (SAC) process. SAC enables the transistor's metal contacts and polysilicon gate to be closer to each other without creating an electrical short circuit.

Some companies offer a mix of 6T and 4T2R SRAMs in their product lines, with 6T cells for low-power devices and 4T2R for more mainstream memories. However, even companies offering no low-power SRAMs plan a conversion to scalable 6T SRAM structures at or near the 0.25- μ m process generation in most cases. Such a conversion will also more easily allow companies to produce logic, memory, and mixed logic/memory chips on the same manufacturing lines.

HOW LOW CAN THEY GO?

One common low-power SRAM design technique, address-transition detection (ATD), involves powering off entire sections of the chip at the conclusion of a read or write until the memory detects the initiation of the next

access cycle. Low-power-SRAM suppliers also adjust internal-transistor switching thresholds to minimize leakage current and don't pump internal bit lines and word lines to halfway between 1 and 0 levels between accesses. Voltage pumping minimizes signal rise/fall times, improving de-

vice performance but at the expense of much higher power consumption, because the pumps aren't 100% efficient and the technique holds CMOS transistors in their high-current linear operating regions.

Reducing standby-current draw is only part of the low-power equation. To reduce active current consumption, vendors carefully lay out their SRAM chip designs,

AFTER MINIMIZING CURRENT DRAW, THE NEXT STEP TO REDUCING POWER CONSUMPTION IS TO LOWER THE OPERATING VOLTAGE.

minimizing trace lengths and removing unnecessary parasitic capacitance. They again trade off performance for power by reducing output-buffer strength and switching speed and by detuning sense amplifiers. Subdividing the entire memory array into multiple smaller regions increases the required decoding logic and therefore the total die size, but it also reduces the percentage of the chip that must power up to access a given address's data (Figure 4).

After minimizing current draw, the next step to reducing power consumption is to lower the operating voltage. Nowadays, 3.3V SRAMs are common, and less-than-3V memories are almost as prevalent. A few companies even sell devices operating at voltages lower than 2V. For example, Enable Semiconductor offers 1.5V extended-temperature, 1- and 2-Mbit, $\times 8$ SRAMs and will soon begin making $\times 16$ devices available for sampling. These memories also support an optional page-access mode, which further reduces power consumption by minimizing the number of addresses you need to toggle for consecutive accesses. To some degree, lower voltage is a natural outcome of advanced manufacturing processes, which create smaller transistors that can no longer tolerate 5V.

If you want to squeeze maximum battery life from your portable system and you have control over its operating voltage,

consider an SRAM with ultra-low-voltage data retention. Reducing the voltage to 1.2V or so gets your device to a few hundred nanoamps of current draw, and you can ramp the voltage back up to the normal operating range when necessary, such as in response to a user key press. Low, reliable data-retention voltages reveal another strength of 6T cell SRAMs over 4T2R alternatives. However, you have to determine whether the other chips on your board can operate at this low level. If not, you need to segment the system design into multiple voltage planes.

Unfortunately, portable communications and computing systems also tend to require wide operating-temperature ranges, which can run counter to your aggressive speed and power-consumption goals. Either negotiate some intelligent compromises with your SRAM vendors to meet the specifications you feel are most important, or brace yourself to pay a lot more money because of the low testing yield the manufacturers achieve because of your constraining parameter goals.

High-performance SRAMs with 32-bit-wide and wider buses and the resulting large, high-pin-count packages are motivating vendors to consider smaller alternatives, such as one of the many BGA options now appearing (Reference 4). Even when portable electronics use narrower 8- or 16-bit SRAMs, a BGA or another chip-sized package (CSP) offers welcome board-space savings. CSPs deliver another advantage: lower lead inductance and capacitance than QFPs, TSOPs, and other traditional packages. This advantage can improve SRAM performance without resorting to difficult-to-handle bare-die alternatives. \square

References

1. Application note "ZBT SRAMs: system design issues and bus timings," Integrated Device Technology Inc.
2. Dipert, Brian, "Advanced DRAM puts you in the fast lane," *EDN*, Oct 9, 1997, pg 52.
3. Lipman, Jim, "How copper-topped silicon creates faster, longer-lasting chips," *EDN*, June 4, 1998, pg 48.
4. Lipman, Jim, "New IC packages really pack in the leads," *EDN*, Sept 1, 1997, pg 93.
5. "Synchronous SRAMs for RISC" (Application Note), *IBM MicroNews*, Second Quarter 1997, pg 29.

You can reach
Technical Editor
Brian Dipert at
1-916-454-5242, fax 1-
916-454-5101, e-mail
edndipert@worldnet.
att.net, URL [http://
members.aol.
com/bdipert](http://members.aol.com/bdipert).

