# SILICON MAGIC: DVINE-LY INSPIRED?

## New Architecture Merges Vector Engines, Embedded DRAM

### By Peter N. Glaskowsky {3/27/00-02}

There's a new entrant in the already crowded field of chip multiprocessor (CMP) media processors. Silicon Magic's DVine architecture, named after its combination of embedded DRAM and vector engines, is most similar to Dave Patterson's V-IRAM design (see

MPR 3/9/98–04, "New Processor Paradigm: V-IRAM") and Cradle's Universal Microsystem (UMS; see MPR 10/6/99-05, "Cradle Chip Does Anything"). DVine is simpler and narrower in focus than most media processors, which are often meant to cover a wide range of potential applications. DVine has a single purpose: to replace fixed-function audio and video codecs in consumer electronics. Silicon Magic says its approach will increase performance and flexibility while cutting development time and chip count.

Patterson's V-IRAM, also based on a symmetric array of vector processors integrated on a single chip with embedded DRAM, is a much simpler design—more like an intelligent RAM, as its name implies. Where the V-IRAM has one scalar RISC processor driving multiple vector engines—and no peripherals—the DVine architecture consists of multiple compute modules (CMs) each with one RISC engine and one vector engine. To these CMs, DVine adds multiple independent memory interface units (MIUs), each driving multiple banks of embedded DRAM. Each MIU includes a simple programmable engine designed to perform basic data-manipulation tasks such as interleaving, interpolation, and rounding. CMs, MIUs, and an external 64-bit bus interface are interconnected via multiple 128-bit buses managed by an on-chip data-flow controller (DFC).

DVine also shares many characteristics with Cradle's UMS, another chip-multiprocessor architecture. UMS offers finer-grain parallelism, with each processing node (or

"quad") combining four simpler 32-bit RISC engines and eight digital-signal engines (DSEs). Cradle's design also includes programmable protocol engines to drive reconfigurable I/O circuitry and an on-chip DRAM controller for external SDRAM.

(Though Improv System's Jazz PSA (see MPR 3/27/00-03, "Jazz Joins VLIW Juggernaut") is also based on a chip-multiprocessor architecture, it reflects different design goals and requires a very different development process.)

Because of Silicon Magic's more limited focus, DVine lacks integer multiplier units and has no support for floating-point calculations. V-IRAM and UMS offer both capabilities, as do most other media processors, making them better choices for applications requiring the dynamic range of floating-point data or the signal-processing throughput enabled by single-cycle multiply-accumulate units.

For some types of audio and video processing, especially those based on MPEG's discrete cosine transform (DCT) and motion-detection algorithms, DVine offers good price/performance. Each of DVine's CMs requires just over one million transistors of logic (including caches and SRAM), about a third as many as are used in each of Cradle's quads (which have significantly more local SRAM storage). In a 0.25-micron process, Cradle's larger and faster (320MHz) quads are expected to deliver 6 GOPS (billion operations per second) on 32-bit integer calculations, edging out DVine's 166MHz and peak throughput of 5.3 GOPS per CM on 16-bit data in the same process. The difference

in clock speeds is primarily attributable to Silicon Magic's use of an embedded-DRAM process that yields slower logic transistors than the logic-only process used by Cradle. Another significant difference between the two competing products is that Silicon Magic is already manufacturing a reference chip; Cradle has not yet begun production of its chip. Silicon Magic expects DVine to reach 200MHz in 0.18-micron implementations.

## Chip Multiprocessing Permits High Performance

DVine's strength is its 16-way "V16" vector engine, able to perform up to 32 parallel 16-bit integer operations per clock cycle. This peak rate is achieved on complex functions such as the sum of absolute differences. The V16 has no local register set; instead, it uses 2K (four 512-byte banks) of local SRAM for local operand storage. The V16's instruction set includes looping constructs but no conditional or branch instructions. Without a hardware multiplier, multiply-accumulate operations take nine clock cycles for each set of 16 results. Each V16 is implemented in about 100K gates of logic, not counting its associated SRAM.

The RISC engine ("REX") in each node adds further performance and flexibility. Silicon Magic says the REX is a simplified version of the DLX machine described by Patterson and John Hennessy in their book *Computer Architecture: A Qualitative Approach*, but with a few added instructions, such as Huffman table lookup and new branch types. The REX core uses a five-stage single-issue 32-bit pipeline with in-order execution and just 30K gates. Silicon Magic says a 200MHz REX engine delivers 167 mips of sustained

performance on representative code (MPEG-2 encoding and decoding algorithms). The REX has a 2K two-way set-associative instruction cache, a set of 32 general-purpose 32-bit registers, and another set of 32 control registers for interrupts, DMA control, V16 handshaking, timers, and so on. The REX core also has two 512-byte blocks of local SRAM storage.

The REX core is also responsible for managing data and instructions for its associated V16. Vector instructions are stored in two 512-byte SRAMs available to the REX and V16. This SRAM, and the other SRAM blocks in the compute module, may also be accessed by a DMA controller in each CM. The DMA engine is used to move data between SRAM and DRAM, where it can access rectangular subsets of two-dimensional arrays—a useful feature for managing raster-scan video frame buffers.

Cradle's UMS, in comparison, provides less local data storage for its DSEs (96 32-bit words) but has a comparable amount of local program storage (384 20-bit instructions). Each UMS quad also has 16K of data memory and 12K of program memory for its RISC cores, much more than is available within DVine's CMs.

Figure 1 shows Silicon Magic's first DVine-based chip, a reference implementation for software development. The 144mm$^2$, 0.25-micron chip is configured with 4M of embedded DRAM and six compute modules. Silicon Magic expects the same chip design in 0.18-micron technology to require just 80mm$^2$ of silicon for the whole chip, 6mm$^2$ per CM. Power consumption for the 0.18-micron implementation is estimated at just 0.3mW/MHz for each CM (0.36W for all six CMs at 200MHz) By comparison, each node in the Cradle UMS is about 13.5mm$^2$ in a 0.25-micron pure-logic process.

## Embedded DRAM Runs Fast on Little Power

Silicon Magic notes that its embedded DRAM is faster than discrete DRAM, consumes less power, and reduces the number of packages and pins in the system. It's also easier to configure a chip with a specific amount of embedded DRAM; if the application needs just 2.5M, the chip can be built with exactly 2.5M. These are definite advantages for portable consumer electronics—one market Silicon Magic hopes to crack with the new architecture. Most such products need a relatively modest amount of DRAM, relying on flash memory or disk drives for mass storage.

This targeting may simply be a consequence of Silicon Magic's
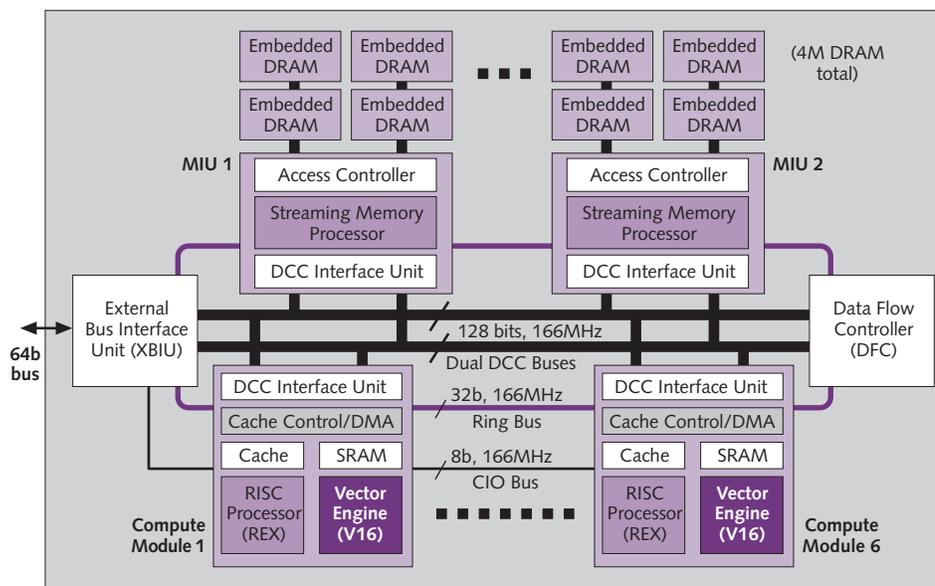


**Figure 1.** The first DVine-based chip, used in Silicon Magic's DVine development system, integrates 4M of embedded DRAM with six compute modules. This chip is 144mm$^2$ in a 0.25-micron process. DVine's scalable architecture allows the number of compute and memory modules to be matched to the needs of a specific application.

expertise in embedded DRAM, however. As the saying goes, "If all you have is a hammer, every problem looks like a nail." Though Silicon Magic has considerable expertise with embedded DRAM, its focus on this technology imposes significant limits on the range of applications DVine can cover. Without an off-chip DRAM controller, DVine is appropriate only for tasks that fit into the available embedded DRAM—up to about 16M in today's process technology—and the higher cost per bit of embedded DRAM, compared with that of discrete DRAM, may dissuade some potential customers. The lower density and slower logic speed of the DRAM process will also increase the effective cost of the DVine processing nodes; each will be larger, and more may be needed to achieve the required throughput.

Without embedded DRAM, DVine chips could be substantially smaller and faster. In some applications—including many audio- and video-related products such as MP3 audio and DVD video players—a single commodity DRAM chip would provide more than enough bandwidth at a lower cost per bit than Silicon Magic's embedded DRAM. Some of today's most popular laptop-PC graphics chips use an intermediate solution—low-cost multichip modules combining the graphics chip with one or two discrete DRAM chips. The power consumption, price, and performance for this approach fall between those of the embedded- and discrete-DRAM alternatives.

For applications that need more bandwidth than is available from discrete DRAMs, the DVine approach can be very effective. Each memory interface unit (MIU) includes a programmable streaming-memory processor (SMP). The SMP operates on data moving in or out of DRAM and performs functions that include interleaving and deinterleaving, decimation, interpolation (useful when moving data from one frame buffer to another with a different resolution), and rounding. In many cases, the SMP will reduce the amount of data being moved around within the chip, effectively increasing on-chip bandwidth. Each MIU is $3mm^2$ in size and operates on 9mW/MHz.

Each MIU connects to two DRAM modules, each of which consists of two DRAM banks in the initial implementation. The DRAM banks themselves are 128 bits wide and operate at the core clock rate, achieving a peak bandwidth of 3.2GB/s at 200MHz. The MIU may be configured to access each module individually to reduce power consumption, or to ping-pong between modules to hide precharge delays and improve sustained throughput. The best-case read-access delay from the REX to any memory bank is six clock cycles (just 30ns at 200MHz), though arbitration and DRAM column-access delays can add extra cycles of delay.

On-chip data transfers are conducted over multiple 128-bit data-communications-channel (DCC) buses that run at the speed of the chip's processor cores. These buses are connected to all on-chip resources. These transfers are managed by a data-flow controller (DFC) unit that also includes other central resources, such as timers and semaphore

registers. With 6.4GB/s of total bandwidth at 200MHz, the DCC buses are unlikely to be a bottleneck for most DVine implementations. Though the prototype chip has two DCC buses, Silicon Magic says it can configure chips with more than two channels—up to as many as one per MIU—if more bandwidth is required.

A 32-bit ring bus provides an additional channel for intrachip communications. The core-speed ring bus is used to pass control and communications information between DVine modules. Any CM or MIU can use the ring bus to query or modify registers in other CMs and MIUs. These transfers are used to request and acknowledge data transfers on the primary data buses.

An 8-bit configurable I/O (CIO) bus connects all CMs to eight pins managed by the external bus interface unit (XBIU). The XBIU also connects an external 64-bit bus to the on-chip DCCs. This bus is DVine's primary interface to external peripherals and a host processor; it operates at 54MHz to achieve 432MB/s of peak throughput.

Silicon Magic expects that many DVine implementations will use integrated peripherals rather than external peripheral chips. Integrated peripherals may be connected to the DCCs, ring bus, CIO bus, or a combination of these. A controller for DRAM expansion could be integrated in this way, for example.

## Simple Software Development Tools Provided

Silicon Magic offers an integrated development environment with a project manager, compilers, simulators, source-level debugger, execution-trace viewer, and optimization tools. Software development is C-based, but Silicon Magic does not have a vectorizing C compiler; most code for the vector engines must be written in their native machine language. Silicon Magic will provide its customers with optimized code libraries for the most popular algorithms, including the elements necessary to implement MP3 and DVD players, digital video recorders, and similar products.

Silicon Magic's tool chain is similar to those available for most media processors. DVine's vector-processor approach is easier to understand and thus more amenable to hand-tuned coding than VLIW architectures, but VLIW designs with good compiler technology, such as Equator's, lead to superior programmer productivity.

Silicon Magic offers no real-time or embedded OSs for DVine at this time. For most purposes, a DVine processor must be teamed with a conventional embedded microprocessor to host the operating system and user-interface tasks.

Silicon Magic touts the scalability of its new architecture, and it's easy to see that DVine can be implemented with just one compute module—or dozens. Whether or not this flexibility will prove useful to potential customers will depend heavily on the target application. There is little support in the DVine architecture for distributing one task across multiple CMs, and no support at all for sharing one CM among multiple tasks. Few real-world tasks will fit

neatly into one CM, and any mismatch will result in some wasted capacity.

The DVine hardware architecture is unusual and severely limited in some ways (such as the lack of support for MACs, floating-point, and fast external memory). These constraints will complicate software development and limit the effective scalability of the architecture. Silicon Magic is considering various extensions to the DVine architecture, but the company is not yet ready to discuss a timetable for these improvements.

With three different programmable elements (REX, V16, and MIU), each with its own programming language, it will be difficult for most customers to reach the full potential of the DVine architecture without help from Silicon Magic. This is a problem shared with all announced CMP media processors, however. Some, particularly Cradle's UMS and Improv's Jazz, have even more complicated development environments. All of these companies hope their benefits—performance, flexibility, and ease of design reuse—offset their higher initial development cost.

For some purposes, the DVine architecture may provide the perfect combination of features. Silicon Magic says a DVD decoder may be implemented with just two CMs and 2M of embedded DRAM in a chip that should cost only a few dollars to manufacture in high volume. Adding incremental features to such a design can be as simple as designing a new chip with more CMs and more memory and writing the necessary code.

There are many other ways to build DVD players, however, and some are no more expensive than the DVine approach. Even if the development cost of the ASIC solution is slightly higher, the lower manufacturing cost of an ASIC will save money in the long run for products manufactured in high volume. Silicon Magic's DVine may need divine guidance to succeed in a market already full of media processors. ◇