G4 Is First PowerPC With AltiVec

Due Mid-1999, Motorola's Next Chip Aims at Macintosh, Networking

by Linley Gwennap

Motorola will extend its PowerPC line with the first G4 processor core, which it unveiled at last month's Microprocessor Forum. According to project leader Paul Reed, the new core adds a faster floating-point unit to the older G3 core and doubles its cache and system-bus bandwidth. The G4 is also the first chip to incorporate the AltiVec extensions, which greatly increase performance on many bandwidth- or compute-intensive applications, including both integer and floating-point algorithms. The chip is already sampling and is due to appear in products by mid-1999.

After IBM pulled out of Somerset (see MPR 6/22/98, p. 4), the G4 became the sole property of Motorola; IBM has no plans to sell the part. Motorola will supply the G4 to Apple for use in its Macintosh products, but the chip vendor also expects the new CPU to be popular in embedded applications that demand high performance and high bandwidth. These applications include network routers, cellular base stations, and cable modems.

FP, Multimedia Performance Improved

The G4 starts with the same integer core as the G3 (see MPR 2/17/97, p. 10), which in turn is not that much different from the original PowerPC 603. The integer core uses a simple five-stage pipeline (four stages for ALU operations) with a limited amount of instruction reordering. Each cycle, the CPU can issue two instructions plus a branch. As Figure 1 shows, integer execution resources include dual ALUs plus a "system" unit (which handles the more complicated integer instructions) and a load/store unit.

Although the integer core was deemed acceptable, Motorola felt the floating-point unit needed improvement. For single-precision instructions, the G3 FPU is fully pipelined with a three-cycle latency, but it requires an extra cycle for double-precision operations, halving the issue rate and increasing the latency. The G4 FPU, however, is fully pipelined, even for double-precision operations. Most technical FP applications use double precision and will see a large speedup on the G4, although single-precision code, such as in 3D games, will see little gain.

The biggest change from the G3 design is the AltiVec unit. This unit implements Motorola's new vector instructions (see MPR 5/11/98, p. 1) and includes its own register file with 32 entries of 128 bits each. Each register can hold up to four FP values or up to 16 short integers.

The G4 can issue two AltiVec instructions per cycle, as Figure 2 shows. All instructions are fully pipelined. The vector FPU has four single-precision multiply-add units, each with a four-cycle latency, that operate in SIMD fashion. Using MAC instructions, peak performance at 400 MHz is 3.2 GFLOPS. A separate permute unit handles AltiVec's powerful permute instruction as well as simpler shifts.

AltiVec also includes SIMD integer instructions, which are similar to Intel's MMX instructions but handle twice as much data. These integer instructions are routed to either the vector simple unit or, for more complex instructions such as multiply-accumulate, to the aptly named vector

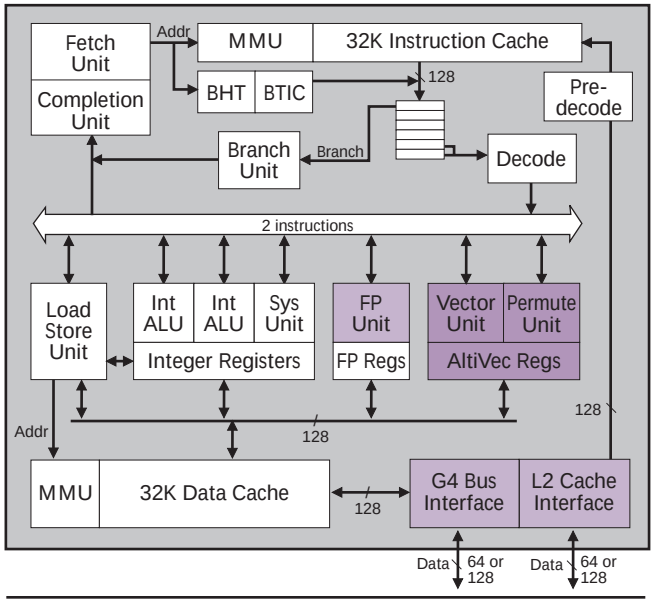


Figure 1. The G4 is based on the earlier G3 processor but includes a revised floating-point unit, a new AltiVec unit, and modified bus and L2 cache interfaces. Internal bandwidth is also improved with a new 128-bit load/store bus.

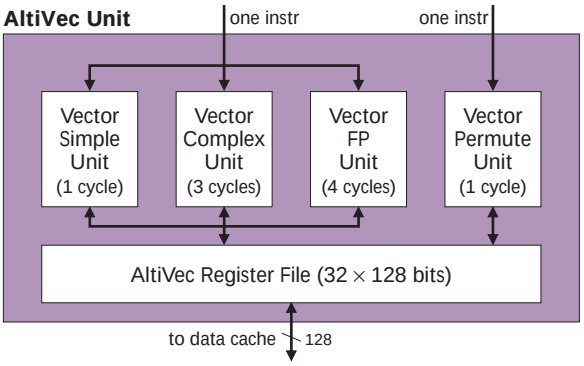


Figure 2. AltiVec instructions are processed by separate units, depending on whether they require integer or floating-point computation.

Price & Availability

Motorola says processors using the G4 core will be available by mid-1999. The company has not announced specific products or prices. For more information on the G4, access *www.moto.com/powerpc*.

complex unit. These instructions have a three-cycle latency versus just one cycle for the simple integer instructions.

The AltiVec unit provides huge performance increases on applications that can take advantage of its SIMD instructions (and have been recoded to do so). On applications such as FIR filters, image processing, and encryption, Motorola has measured speedups of 9× to 15× over the same applications, using standard PowerPC instructions.

Better Bandwidth Boosts Performance

The bottleneck for many applications, particularly in the multimedia and networking areas that Motorola is focused on, is system bandwidth rather than integer performance. As a result, Reed's team reworked the bus and memory interfaces to improve bandwidth in several areas. Starting inside the chip itself, we see the internal memory bus is 128 bits wide, twice as wide as in the G3. This allows an entire AltiVec register to be loaded or stored in a single cycle.

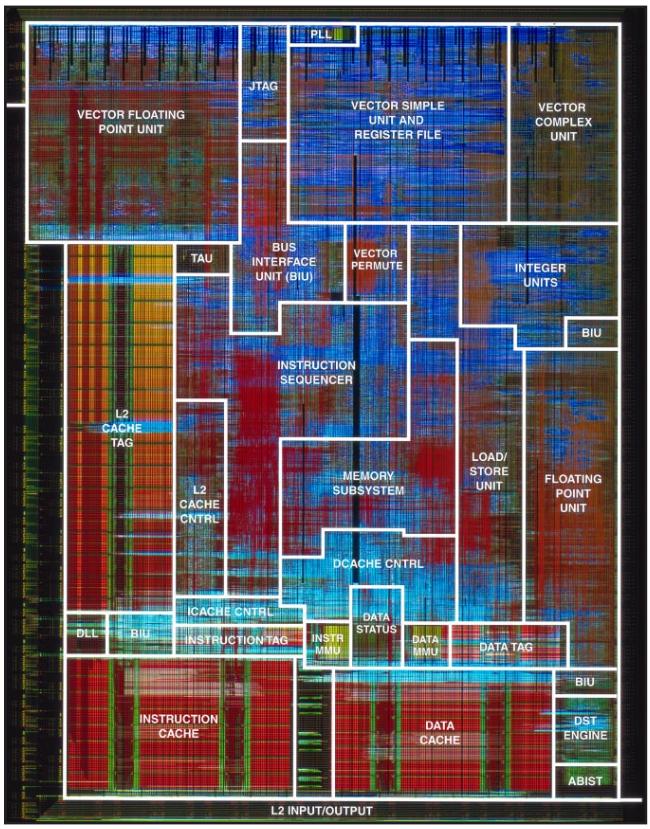


Figure 3. Motorola's first G4 processor is 8.25 mm by 10.1 mm when fabricated in a 0.22-micron five-layer-metal copper process.

The connection to the external level-two (L2) cache is also extended to 128 bits, although the chip retains a 64-bit option for low-cost systems and for compatibility with the G3-based PowerPC 750. When the wider bus is used, a complete 32-byte cache line can be moved from the fully pipelined L2 cache to the L1 in just two cycles.

Like the 750, the new processor includes the L2 cache tags on the chip, but the G4 doubles the maximum supported cache size to 2M. While the 750 maintains the L1 data cache as a subset of the L2 cache, the G4 moves requested data into the L1 cache only. This method reduces L2 cache traffic, in particular when data is modified in the L1 cache before being evicted to the L2.

The L2 cache is nonblocking, meaning it can continue processing requests even while a cache miss is pending. Unlike many nonblocking caches, the G4's cache can have multiple pending cache misses, up to eight.

The G4 supports a variety of SRAM types and speeds, including caches at the full CPU speed. Most designs, however, will use a 2/3- or 1/2-speed cache for cost reasons. The latency to the L2 cache is 8 CPU cycles for a full-speed design or 11 cycles for a half-speed cache.

Taking advantage of the data-stream touch feature of AltiVec, the G4 includes four autonomous prefetch engines. As directed by DST instructions, these engines will load data into either the L1 or L2, using otherwise empty bus cycles. In cases where the CPU is processing a large array of data, these prefetch engines can nearly eliminate memory latency without consuming CPU resources.

Faster Bus Improves MP Support

The G4 bus uses a superset of the established 60x bus protocols and can handle either 64 or 128 bits of data. The G4 can drop into existing 60x-bus designs using a fully compatible 64-bit mode. For better performance, new systems can use the enhanced protocols with a 64-bit bus or a 128-bit bus.

While the 60x bus already handles split transactions, the enhanced protocols support out-of-order transactions, with up to seven outstanding transactions per processor. In multiprocessor (MP) systems, data can be copied directly from one processor's cache to another's. Other minor changes reduce the number of dead cycles on the bus in certain situations.

Although G3 processors can be used in MP systems, they require external support logic to connect more than two chips. The G4 is designed for glueless multiprocessing with up to four processors. Unlike the G3, the new chip implements the full MESI protocol and even adds a fifth "reserved" state to enable processor-to-processor transfers on shared data. Dual-ported tags on the data cache allow snoop transactions to occur without slowing down the CPU.

The G4 bus also enables higher clock speeds. The 60x bus operates with 3.3-V signals and is limited to 100 MHz

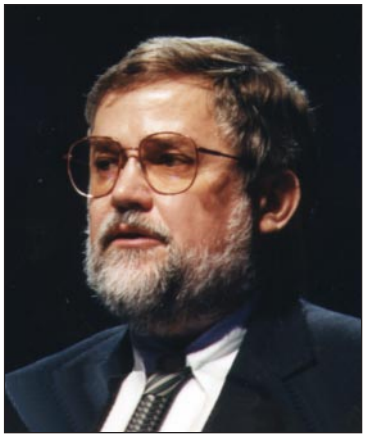
at best; the G4 uses 1.8-V signal levels to simplify 100-MHz operation. Motorola expects the new bus to scale "well beyond" 100 MHz in the future.

In a G4 system with a 400-MHz CPU, a 200-MHz L2 cache, and a 100-MHz system bus, the L2 cache delivers an impressive 3.2 Gbytes/s, while the system bus reaches 1.6 Gbytes/s, assuming both ports operate at the full 128-bit width. These figures are twice the bandwidth achieved in a Pentium II, which uses 64-bit interfaces for both of its ports.

Strong Performance From Small Die

The G4 is built on Motorola's HIP5 (see MPR 9/14/98, p. 1), a 0.22-micron process with six copper metal layers. In this process, the chip is expected to achieve clock speeds of at least 400 MHz; exact speed grades will be set at the product introduction next spring. Reed reported that the chip uses 10.5 million transistors and measures 83 mm^2 in this process.

The G4 is more than twice the size of IBM's Lonestar (see MPR 9/14/98, p. 4), a G3 processor built in a nearly identical process. As Figure 3 shows, the AltiVec units provide much of the increase, consuming about 20% of the new die. The 64-bit floating-point unit, wider buses, and the doubling of the L2 cache tags also contribute to the increase in size.



MICHAEL MUSTACCHI

Motorola project leader Paul Reed describes the G4's performance-enhancing AltiVec unit.

Even at 83 mm^2, the G4 is still much smaller than Intel's Pentium II processors; the Deschutes version, for example, measures 118 mm^2 in Intel's latest P856.5 process. The MDR Cost Model estimates that the G4 costs $45 to build, versus $55 for a Deschutes CPU (not including the Pentium II module and L2 cache costs). Lonestar, in contrast, costs just $30 to build. The relatively low manufacturing cost of the G4 will allow Motorola to offer aggressive pricing to gain design wins in networking and other high-end embedded applications.

In 1H00, Motorola is likely to move the G4 into its 0.18-micron HIP6 process. This move will reduce the chip's die size to about 50 mm^2 and the manufacturing cost to about $35. The new process should also increase clock speeds to well in excess of 500 MHz.

The initial versions of the G4 will use the same 360-pin BGA package as the PowerPC 750 (G3), allowing the new chip to drop into existing system designs. This package, however, will not support the wider cache and bus interfaces. Motorola plans to deploy other versions that enable the higher-bandwidth buses but did not provide details.

Despite its larger die and higher speed, the G4 dissipates less than 8 W, enabling it to be used in mobile computers. The new chip takes advantage of the 1.8-V operation of the HIP5 process; the use of copper also reduces power dissipation. The future 0.18-micron version will further reduce power dissipation at a given clock speed.

Motorola expects to use the G4 core in several products with feature sets tuned for different markets. For example, the FPU could be omitted for some cost-sensitive embedded applications. A version specifically for networking might include a slew of communications channels. The company is also hard at work on a completely new core, sequentially denoted as the G5, that is slated to appear in 2000.

G4 Boosts Macintosh Against Katmai

The G4 will help Apple compete against Katmai-based PCs in 1999. AltiVec should deliver better performance than the Katmai New Instructions (see MPR 10/5/98, p. 1) on many multimedia applications. At an estimated 18 SPECint95 and 18 SPECfp95 (base), a 400-MHz G4 may be slightly behind a Katmai-500 on the integer side but well ahead in floating-point performance. In the latter area, the G4 gains due to its fully pipelined FP multiply-add unit; Katmai's FPU can issue only one multiply-add every two cycles.

The problem for Motorola is that Intel will move Katmai into a 0.18-micron process in 3Q99, shortly after the G4 debuts. In this process, the Katmai core should quickly reach 600 MHz, with 733-MHz parts likely in 1H00. Even when Motorola moves to its own 0.18-micron process, it won't be able to push the clock speed of the G4 as high, due to its stubby five-stage pipeline; Katmai, in contrast, is built for speed, with a much deeper pipeline (even after excluding x86-specific stages).

Thus, the Intel line is likely to retain an edge on standard integer applications, but the G4 will be a better solution for many floating-point and multimedia applications. But the two competing processors will be close enough in performance that anyone choosing between a PC and a Macintosh is unlikely to be swayed by the differences in the CPU.

In the embedded market, the G4 will set a new standard of performance, outrunning the 300-MHz R7000 and other fast low-end chips. Its AltiVec unit will outmuscle Hitachi's SH7750 on 3D graphics and could allow the G4 to replace an entire bank of DSP chips in a cellular base station or other compute-intensive environment. The G4's memory bandwidth is unprecedented in the embedded market, making the chip a natural for high-end networking devices. At 400 MHz, however, the G4 could be quite pricey.

The G4 is likely to be most successful initially in the Macintosh line, pushing IBM out of that account over time. Even before Apple can ship G4 products, the new chip may appear in a few performance-intensive embedded designs that can afford a premium CPU. To see widespread embedded use, however, the G4 will probably need a shrink to 0.18-micron technology to reduce its cost. That progression should ensure a long life for the G4 at Motorola. M