

# MICROPROCESSOR REPORT

THE INSIDERS' GUIDE TO MICROPROCESSOR HARDWARE

## PC Processors for Model Year 2000

### A Comparison of Microarchitectures in Upcoming PC Processors

by Keith Diefendorff

There is no objective way to ascertain which is the "best" microprocessor for a PC. Which attributes are most important is unclear, and how best to measure their value is debatable. Benchmarks are valuable tools, but the results are often misused or misunderstood, and they can obscure important characteristics. Moreover, benchmarks are not available for processors that haven't yet reached the market. For chip suppliers this is a good thing; without the cover of confusion and uncertainty, many of them wouldn't survive.

With no illusion of picking a winner, but hoping to shed some light on the subject, we review the microarchitectures of the upcoming crop of PC processors, including Intel's Pentium III (Coppermine), AMD's Athlon, Cyrix's Mojave, Centaur's WinChip 4, Rise's mP6 II, and Motorola's G4. As background, in the previous issue (see MPR 7/12/99, p. 16) we presented a concise overview of the microarchitecture techniques used in modern PC processors.

#### Intel, AMD Vie for Performance Lead

Intel will introduce its forthcoming Coppermine version of Pentium III at the high end of the market and, as it has done with many previous processors, drive it aggressively down the price curve until it becomes a low-end processor. We expect Coppermine, or simple derivatives of it, to enter the low-end market segments within about six months of introduction.

Coppermine is a 0.18-micron version of the 0.25-micron Pentium III (Katmai). Pentium III, like Pentium II before it, is based on the P6 microarchitecture, which Intel first introduced in Pentium Pro in 1995. Coppermine takes advantage of Intel's new P858 process to boost frequency and bring 256K of L2 cache onto the chip. Intel says Coppermine will enter production this November. We expect it to begin life at up to 667 MHz, as Figure 1 shows.

Weary of competing against Intel with the K6, AMD has quietly begun sampling Athlon. Athlon is based on AMD's new K7 microarchitecture—the most powerful of

any ever deployed in a PC processor. Unlike the K6, the K7 uses a long pipeline, designed to enable AMD to match Intel on the all-important clock-frequency parameter. We anticipate that AMD will introduce Athlon at up to 650 MHz in 0.25-micron CS44E this quarter. AMD says it will ship Athlon in 0.18-micron CS50 before the end of the year, not far behind Coppermine, boosting frequency substantially. Assuming AMD can avoid stubbing its toe on the manufacturing line again, Athlon will become the first processor to seriously challenge Intel for the performance high ground.

#### Other x86 Competitors Target Low End

The status of Cyrix's x86 processors is still somewhat uncertain following National's decision to sell Cyrix to Via (see MPR 7/12/99, p. 5). We assume for purposes of this article, however, that Cyrix's roadmap remains intact and, as Cyrix expects, Via will proceed with Mojave.

At Microprocessor Forum last year, Cyrix described the M3—a highly integrated Jalapeno-based processor for which the company now seems to have lost enthusiasm. Mojave uses the same Jalapeno core as the M3, but in a more

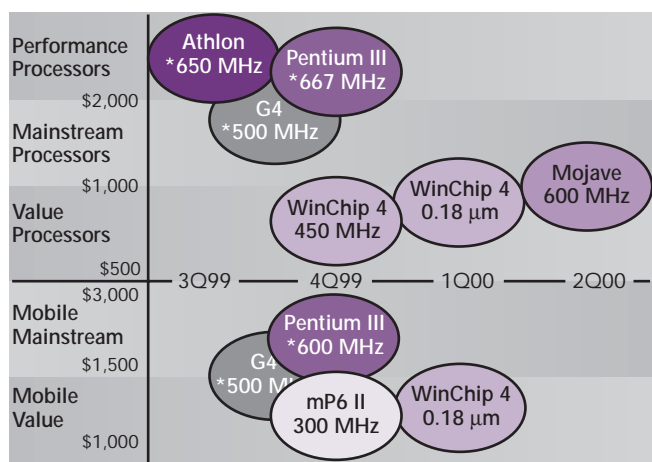


Figure 1. On the basis of their microarchitectures and frequencies at introduction, we expect the new PC microprocessors to be best suited to the segments shown. (\*MDR estimates)

traditional configuration, with a 133-MHz version of Intel's Socket 370 interface. At the Forum, Cyrix claimed the M3 would ship at 600 MHz in National's 0.18-micron CMOS-9 process (see MPR 9/14/98, p. 1); we assume Mojave is targeted at the same speed. (National refers to CMOS-9 as a 0.18-micron process, although our analysis shows it to be more like a 0.21-micron process.)

Like the P6 and the K7, Jalapeno is a superscalar out-of-order design. But unlike those processors, both of which dispatch and decode up to three instructions per cycle, Jalapeno does only two. Its instruction reordering capability is also more limited than that of the P6 or the K7.

Like Mojave, Centaur's WinChip 4 faces an uncertain future, as IDT has decided to exit the x86 processor business and to sell Centaur (see MPR 8/2/99, p. 4). We assume, however, that a buyer will be found and that WinChip 4 will survive. WinChip 4 employs a simple in-order microarchitecture that, unlike previous WinChips, is deeply pipelined to more closely match Intel's frequencies.

Centaur's separation from IDT may be a blessing. The company has had to endure IDT's endless difficulties getting its 0.25-micron CMOS-10.5 process into production, a process Centaur was counting on for low cost while achieving 500-MHz operation. As a result of IDT's difficulties, Centaur has been forced to utilize IBM's 0.28-micron CMOS-6S2 process. This process will probably limit initial WinChip 4 parts to 400–450 MHz and increase costs. It would help Centaur greatly if the new owner has a good, inexpensive IC process.

The latest entrant in the PC processor game is Rise (see MPR 11/16/98, p. 1), a fabless semiconductor startup based in Silicon Valley. Like IDT, Rise has its sights set on the low-

end desktop and portable markets. Its current mP6 processor is a three-issue in-order design, which it will beef up with a 256K on-chip L2 cache in the mP6 II. Rise says the mP6 II will be rendered in an undisclosed 0.18-micron process. Our sources indicate that it may be UMC's L180 process.

The only non-x86 processor family currently in the hunt for a slice of the PC market is PowerPC, which is at the heart of Apple's Macintosh. The next-generation PowerPC processor is Motorola's G4, which uses the same superscalar, out-of-order microarchitecture as the current PPC 750 (G3) but adds Motorola's new AltiVec SIMD architecture. The G4 also includes a fully pipelined scalar floating-point unit and boasts improved memory bandwidth. The chip will initially be implemented in Motorola's 0.22-micron copper HIP5 process. Motorola has not disclosed the initial frequency, but sources indicate it may debut at up to 500 MHz.

### Pipeline and Process Determine Frequency

The first-order terms in the frequency equation are pipeline length and IC process, as Figure 2 shows. Against the standard set by the P6, Athlon easily achieves its expected pipeline-defined clock rate, despite the fact that AMD's CS44E is substantially slower than Intel's P858. This achievement indicates a very evenly partitioned pipeline. Clearly Athlon's pipeline design is superior to the K6's, as Athlon operates at a 25% lower voltage; evidently Athlon doesn't require the transistors to be driven as hard as they were for the K6.

Mojave will debut below its natural pipeline frequencies, probably held there by process-speed limitations. The mP6 II will also operate below its natural pipeline frequency, but it is doing so despite an advanced 0.18-micron process. Perhaps there is a critical speed path that is limiting frequency, or perhaps we have overestimated the process technology Rise is using.

Motorola's G4 is obviously taking advantage of its advanced copper process to operate the G4 above its expected pipeline frequency. The G4, however, like the 750 and the 603 before it, is clearly underpipelined; the design would benefit greatly from a few additional pipeline stages.

### Branch Misprediction Saps Performance

Although long pipelines can increase frequency, the payback comes on conditional branches. The P6 microarchitecture, first deployed in a 0.5-micron technology, was limited in the area that could be invested in dynamic branch-prediction hardware. Although Intel has never fully described the P6's branch hardware, academic studies of similar two-level schemes have shown prediction accuracies in the range of 90% to 95% on several benchmarks.

The newer x86 microarchitectures, initially designed for 0.25- or 0.18-micron processes, have been less constrained by silicon area. The more recent designs have also benefited from recent branch-prediction research and have implemented schemes that are consistently able to achieve

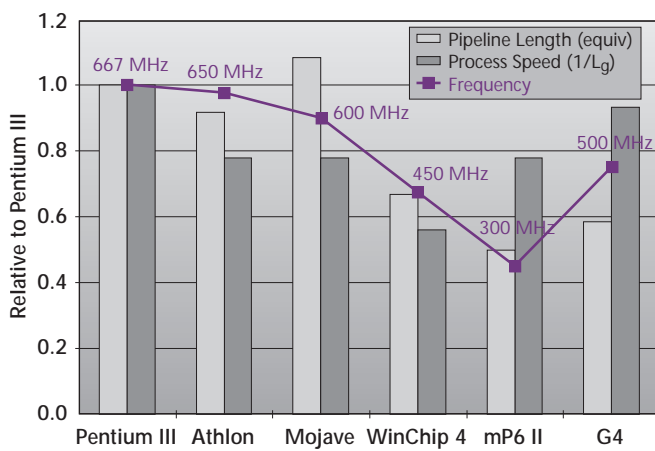


Figure 2. Clock frequency is generally well correlated to the load pipeline length in these processors. The frequency (purple) is our expectation of the maximum frequency of the part at introduction. All values were normalized to Pentium III at 667 MHz. Pipeline length was measured through availability of data from the D-cache. Pipeline length was adjusted by one stage for Athlon and G4 to account for predecode. The G4's pipeline was credited with another two stages to account for its RISC ISA. Process speed was estimated as the reciprocal of the gate length ( $L_g$ ). (Source: MDR)

Feature	Intel Coppermine	AMD Athlon	Cyrix Mojave	IDT WinChip 4	Rise mP6 II	Motorola G4
Reference MPR Article	3/8/99, p. 1	10/26/98, p. 1	11/16/98, p. 24	12/7/98, p. 18	11/16/98, p. 1	11/16/98, p. 17
Scheduled Production	Nov-99	Jul-99	2Q00	4Q99	4Q99	2H99
Microarchitecture/Core	P6	K7	Jalapeno	C4	mP6	750
Integer Architecture	x86 (CISC)	x86 (CISC)	x86 (CISC)	x86 (CISC)	x86 (CISC)	PowerPC (RISC)
Floating-Point Architecture	x87 (stack)	x87 (stack)	x87 (stack)	x87 (stack)	x87 (stack)	PowerPC (flat)
SIMD Int/FP Architecture	MMX/SSE	MMX/3DNow	MMX/3DNow	MMX/3DNow	MMX/None	AltiVec/AltiVec
GP Registers, FP Registers	8 × 32, 8 × 80	8 × 32, 8 × 80	8 × 32, 8 × 80	8 × 32, 8 × 80	8 × 32, 8 × 80	32 × 32, 32 × 64
SIMD Integer Vector Regs	Uses FPRs	Uses FPRs	Uses FPRs	Uses FPRs	Uses FPRs	32 × 128
SIMD FP Vector Regs	8 × 128	Uses FPRs	Uses FPRs	Uses FPRs	None	Uses SIMD IVRs
Instruction Predecode	None	3 bits/byte (38%)	None	None	None	4 bits/instr (13%)
Instruction Decode Width	3 x86 (1 + 2)	3 x86	2 x86	1 x86 (2 MMX)	3 x86	3 PowerPC
Dispatch Width	6 ROPs	6 ROPs	3 ROPs	1 x86 (2 MMX)	3 x86	2 RISC + 1 br
Issue Window	20 ROPs	36int/36fp ROPs	6 ROPs	1 x86 (2 MMX)	3 x86	6 RISC
Issue Width	5 ROPs	9 ROPs	6 ROPs	1 x86 (2 MMX)	3 x86	6 RISC
Retirement Width	3 ROPs	6 ROPs	2 ROPs	1 x86 (2 MMX)	3 x86	2 RISC + 1 br
Result Reordering Hardware	Reorder buffer	Future file	Reorder buffer	None	None	Completion buf
Int/FP Rename Registers	40	36/36	64	None	None	6/6
Execution Units (nonbranch)	5 units	9 units	5 units	4 units	7 units	7 units
Int-Ld Pipeline (not incl retire)	10-12 cycles	8-10 cycles	11-15 cycles	11 cycles	8 cycles	4-5 cycles
FP Add (throughput/latency)	1/3	1/4	1/4	2/6	1/4	1/3
FP Mul or <sup>1</sup> Mul-Add	2/3	1/4	1/5	2/6 sp, 4/8 dp	1/4	1/3 <sup>†</sup>
SIMD Add (16b parallelism)	1/1 (×4)	1/2 (×4)	1/1 (×4)	1/1 (×4)	1/1 (×4)	1/1 (×8)
SIMD Mul-Add	1/3	1/2	1/4	1/4	1/2	1/3
FP SIMD Add (SP parallelism)	2/4 (×4)	1/4 (×2)	1/3 (×2)	1/4 (×2)	None	1/4 (×4)
FP SIMD Mul or <sup>1</sup> Mul-Add	2/6	1/4	1/5	1/4	None	1/4 <sup>†</sup>
Branch Predictor	Two-level	GShare	Two-level	Two-level, GShare	One-level	Static + one-level
BHT	512 × 2b*	4,096 × 2b	1,024 × 7b	16K × 1b	512 × 2b	512 × 2b
BTAC or <sup>1</sup> BTIC	512 entries*	4,096 entries	1,024 entries	64 entries	512 entries	64 entry <sup>†</sup>
Mispredict Penalty (min)	10 cycles*	10 cycles	12 cycles	8 cycles	6 cycles	2 cycles
Memory Op Reordering	Ld-St, Ld-Ld	Ld-St, Ld-Ld	Ld-St, Ld-Ld	In order	In order	L-S, L-L, S-S
Store Reservation Stations	12	44	12	0	0	2
Store-Data Forwarding	Yes	Yes	Yes	No	No	No
Nonbinding Prefetch	Yes	Yes	Yes	Yes	No	4 async streams
Instruction TLB ( <sup>†</sup> full assoc)	32 entry <sup>†</sup>	24 <sup>†</sup> + 256, 4-way	32 <sup>†</sup> + 256, 4-way	128, 8-way	32 entry, 4-way	128, 2-way
Data TLB ( <sup>†</sup> full assoc)	64 entry <sup>†</sup>	32 <sup>†</sup> + 256, 4-way	32 <sup>†</sup> + 256, 4-way	128, 8-way	64 entry, 4-way	128, 2-way
L1 Caches (I/D)	16K/16K 4-way	64K/64K 2-way	16K/16K 4-way	64K 2w/64K 4w	8K/8K 2-way	32K/32K 8-way
L1 Banks or Ports	2 banks	2 banks	2 ports	1 port	4 ports	1 port
L1 DC Accesses per Cycle	1 Ld and 1 St	2 Ld or 2 St	1 Ld and 1 St	1 Ld or 1 St	2 Ld and 2 St	1 Ld or 1 St
L1 Access Time	2 cycles	1 cycle	2 cycles	2 cycles	1 cycle	1 cycle
L1 Load Use (to exec/agen)	3/3 cycles	3/3 cycles	4/4 cycles	0/3 cycles	0/2 cycles	2/2 cycles
L2 Cache Location	On chip	External on BSB	On chip	External on FSB	On chip	External on BSB
L2 Tags Location	On chip	On chip	On chip	External	On chip	On chip
L2 Cache Size, Associativity	256K, 2-way*	512K-8M, 2-way	256K, 8-way	512K+, 1-way	256K, 1-way	512K-2M, 2-way
L2 Access Time ( <sup>†</sup> half speed)	6 cycles*	11 cycles <sup>†</sup>	7 cycles	16 cycles	4 (instr), 3 (data)	9 cycles <sup>†</sup>
L2 Cache-Bus Width	256 bits*	64 bits	256 bits	64 bits	256 Bits	64 or 128 bits
Coherency	MESI, snoop	MOESI, snoop	MESI, snoop	MESI, snoop	MESI, snoop	MERSI, snoop
System Bus	P6 bus, Slot 1	EV6 bus, Slot A	P6, Socket 370	Pentium, Socket 7	Pentium, Socket 7	Enhanced 60x
System Data Bus Width	64 bits	64 bits	64 bits	64 bits	64 bits	64 or 128 bits
System Bus Frequency	133 MHz	200 MHz	133 MHz	100 MHz	100 MHz	100 MHz
Transaction Order (normal)	In order	Out of order	In order	In order	In order	Out of order
Split Transactions (CPU limit)	8 pending (4)	24 pending (20)	8 pending (1)	No	No	7 pending (6)
IC Process	Intel P858	AMD CS44E+	National CMOS9	IBM CMOS-6S2	UMC L180*	Motorola HIP5
Lithography	0.18 μm	0.25 μm	0.21 μm	0.28 μm	0.18 μm*	0.22 μm
Gate Length	0.14 μm	0.18 μm	0.18 μm	0.25 μm	0.18 μm*	0.15 μm
Interconnect	6-layer Al	6-layer Al	5-layer Al	6-layer Al	6-layer Al*	6-layer Cu
Max Frequency at Intro	667 MHz*	650 MHz*	600 MHz	450 MHz	300 MHz	500 MHz*
Transistors (core + cache)	8M + 15M	11M + 11M	10M + 15M	4.5M + 7M	3M + 15M	4.5M + 6M
Die Size	103 mm <sup>2</sup>	184 mm <sup>2</sup>	110 mm <sup>2</sup> *	115 mm <sup>2</sup>	105 mm <sup>2</sup>	83 mm <sup>2</sup>
Package	495 OLGA	576 CBGA	370 PPGA	296 SPGA	PBGA on 296 PGA	360 CBGA
Core Voltage	1.5 V*	1.6 V	1.8 V	2.8 V	2.0 V	1.8 V
Power (typ-max)	14-19 W*	34-48 W*	31-43 W*	14-19 W	4.5-6.2 W	7-10 W*

Table 1. A summary of the key parameters of the next generation of PC microprocessors at introduction. See MPR 7/12/99, p. 16 for a definition of the terms used in this table. <sup>†</sup>defined on each line (Source: vendors, except \*MDR estimates)

prediction accuracies in excess of 95%, even on difficult benchmarks like Winstone. The WinChip 4, for example, uses a very complex, yet quite area-efficient, branch predictor that combines several types of predictions, based on the type and recent history of the branch being predicted.

Mojave also implements an aggressive predictor, but it has the longest mispredict penalty of all the processors. The mP6 II's predictor is less aggressive than the others, so, despite its relatively short pipeline, we expect it to have a hefty loop penalty.

The G4 excels on this metric because, even though its branch predictor is only moderately prescient, its mispredict penalty is a mere two cycles. As a result, while the part pays dearly in frequency for its short pipeline, it has excellent branch efficiency. The branch performance of the G4 is actually even better than is indicated in Figure 3. The PowerPC architecture provides multiple condition registers to assist the compiler in statically scheduling conditions. In the frequent case that the compiler can arrange the code to precompute a condition, the G4's branch target instruction cache gives it a zero-cycle mispredict penalty.

### Compute Bandwidth Sets Upper Limit

An often used indicator of a microarchitecture's capability is the instruction throughput it can sustain into an infinite cache. This value is hard to estimate, however, without a simulator and appropriate benchmark code. Therefore, as a less than perfect but somewhat representative measure, we look at the processor's peak compute bandwidth.

This metric can be important in its own right as an indicator of processing-power headroom. Because the ebb and flow of computation is often erratic, a processor that lacks headroom can clip the peak demand, thereby sacrificing throughput. The inner loops of some multimedia and DSP algorithms, for example, can saturate considerable computational resources. Without a high peak compute band-

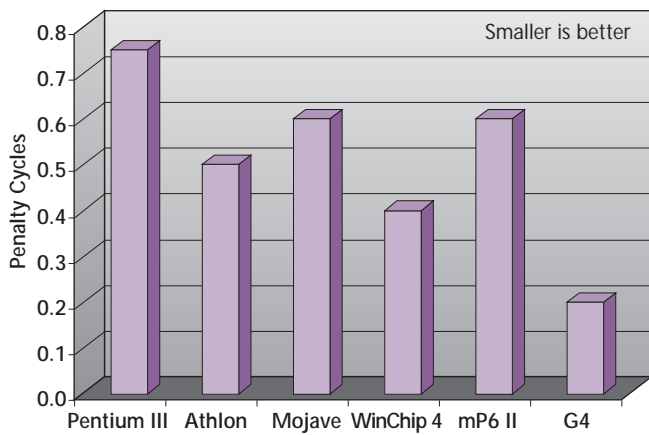


Figure 3. Using the aggressiveness of their branch predictors and the misprediction penalty of their pipelines, our estimates indicate that WinChip 4 and the G4 have the smallest loop penalties (mispredict rate  $\times$  mispredict penalty). The minimum mispredict penalty was used for each chip. (Source: MDR)

width, a processor may not be able to meet the real-time demands of some algorithms, even though it has a very high average throughput capability.

Figure 4 shows that Athlon's three integer units and two fully pipelined floating-point units give it a sizable advantage over Pentium III on scalar code, especially floating-point code. It is also noteworthy that Athlon's 3DNow matches Pentium III's SSE on SIMD-FP throughput, even though, as Table 1 shows, SSE architecturally specifies twice the parallelism. The reason is that Pentium III's SSE execution units, unlike Athlon's 3DNow units, are not fully pipelined.

Mojave's two fully pipelined FPU's also give it a scalar floating-point performance advantage over Pentium III. WinChip 4, on the other hand, falls behind on both scalar integer and floating-point on this metric, due to its single-issue design and its partially pipelined floating-point units.

Fortunately, WinChip 4 can dual-issue MMX and 3DNow instructions, bringing its multimedia capability more on a par with that of the other chips. But the part will have to reach higher frequencies to match the full capability of Pentium III, Athlon, or Mojave. Rise's mP6 II, despite its three-issue capability, suffers on this metric because of its low core frequency. Motorola's G4 has only moderate scalar integer throughput, but it excels at scalar floating-point, thanks to its fully pipelined multiply-add unit, and at SIMD operations, thanks to its two fully pipelined 128-bit AltiVec units.

### Memory Latency Stalls Pipeline

Regardless of peak compute bandwidth, a processor's execution units will sit idle if the processor's memory system can't supply data quickly on demand. Average access time is one indicator of the memory system's ability to deliver data. This parameter determines the average latency of load and store operations. The longer the average latency, the more difficult

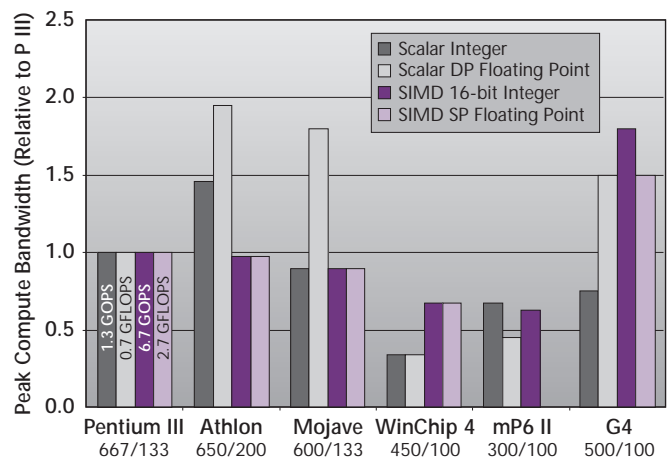


Figure 4. The peak compute bandwidth, i.e., the maximum number of arithmetic operations the processor can sustain out of registers with a perfect instruction schedule, is indicative of the computational power and headroom of a processor. All results are relative to Pentium III at 667 MHz. (Source: MDR)

it is to schedule instructions into the load delay slots and the more stalls that will occur when this effort fails. The calculation we use here for average access time is simplified, ignoring factors such as memory operation reordering and dirty-line copybacks, but it still provides a first approximation of performance of the various memory systems.

Pentium III may actually perform somewhat better than Figure 5 indicates; Intel has not yet released information on Coppermine's L2 cache, so we have made the pessimistic assumption that its access time and associativity are similar to those of Celeron's on-chip L2. The company will probably do better, since Mendocino's L2 was hurriedly moved onto the chip to avert the performance debacle of the initial cacheless Celeron (Covington).

Athlon fares surprisingly well on this measurement, primarily because of its large, L1 caches and its on-chip L2 tags. Mojave also performs admirably, thanks to its highly associative on-chip L2. WinChip 4, however, fares less well, because its L2 cache and tags are off chip, across the slow 100-MHz system bus. The mP6 II's memory hierarchy actually performs pretty well compared with the other chips when measured in CPU cycles. But because the mP6 II's clock rate is so low, the average access time is poor.

The arrangement of WinChip 4's and the mP6 II's pipeline to some extent mitigates the problem depicted in Figure 6. Both chips set the ALU behind the cache in the pipeline, giving them a zero-cycle load-use penalty on L1 hits. The other chips all have multiple-cycle load-use penalties and rely on instruction reordering to cover the load delay slots.

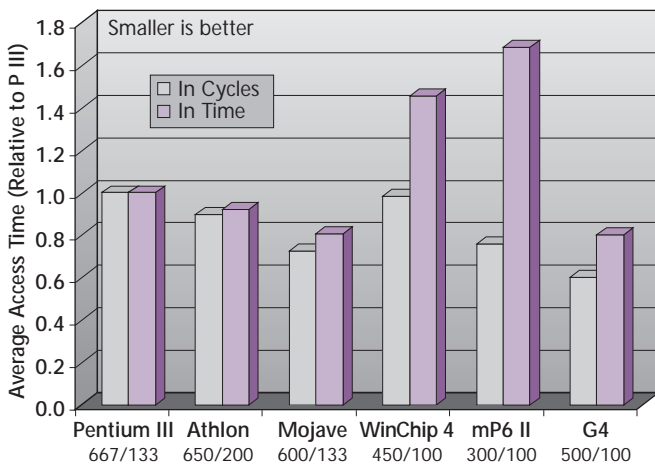


Figure 5. Average access time is one indicator of the performance of a processor's memory hierarchy. For this calculation, we assumed that internal caches run at CPU clock speed; external 512K backside L2 caches run at half the CPU speed; external frontside L2 caches run at the system bus speed; and average DRAM access time is 45 ns plus some delay through the system bus and core logic. Cache miss ratios were assumed to be proportional to the square root of size and associativity. The average access time was computed as  $T_{avg} = T_{accL1} + (Miss_{L1} \times T_{accL2}) + (Miss_{L1} \times Miss_{L2} \times T_{accDRAM})$  and normalized to Pentium III. (Source: MDR)

### Memory Bandwidth Governs Throughput

The other half of the memory-performance equation is bandwidth. Algorithms with high degrees of parallelism and with good spatial locality on instructions and data tend to be sensitive to memory bandwidth. Thus, a processor's memory system must provide high bandwidth as well as low latency.

As Figure 6 shows, the L1 data-cache load bandwidth of Pentium III and Mojave are similar: both have fully pipelined L1s capable of supporting one load and one store per cycle. Mojave actually implements completely independent physical ports, eliminating all bank conflicts. Athlon has the most flexible organization, allowing two 80-bit loads, two 32-bit stores, or one 80-bit load and one 80-bit store every cycle.

The mP6 II achieves good L1 load bandwidth, despite its low frequency, by providing two load and two store ports. In contrast, WinChip 4 and the G4 both have only a single memory port that must be shared between loads and stores, sapping about a third of the peak load bandwidth. The G4's load/store port is 128 bits wide, allowing it to feed the AltiVec register file at full speed. It should be noted that Figure 6 somewhat overstates the magnitude of the L1 bandwidth deficiencies in WinChip 4 and the G4, since the other chips will only occasionally saturate their multiple L1 ports.

Figure 6 also shows that all the chips except WinChip 4 provide similar L2 cache bandwidth. Notice that while the on-chip L2 caches in Coppermine, Mojave, and the mP6 II provide significantly lower latency, they do not generally offer more bandwidth, because, unlike external L2s that are heavily pipelined to get data rapidly across a narrow bus, the wide L2-cache buses available on the processor die make pipelining on-chip L2s unnecessary.

WinChip 4 is penalized severely on L2 bandwidth by the use of its 100-MHz system bus as an L2 cache bus. WinChip 4's large L1s compensate for this deficiency in many cases, but not all. The G4 achieves the highest L2 bandwidth due to its 128-bit backside bus, twice as wide as

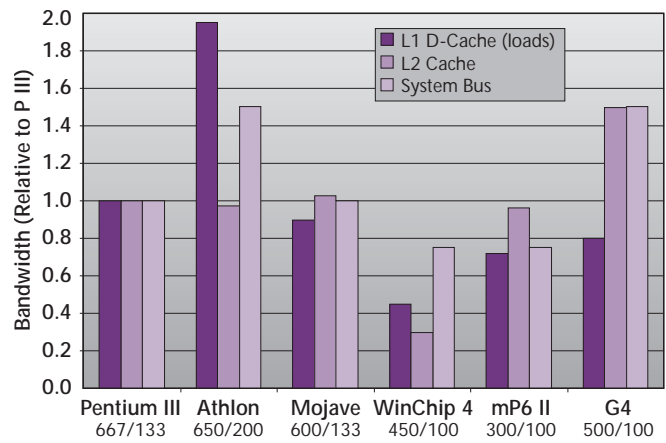


Figure 6. To support high instruction throughput, processors must provide sufficient load bandwidth to avoid starving the execution units. In this chart, we assume that Athlon's and the G4's backside L2s operate at half the CPU speed and that WinChip 4's frontside L2 runs at the bus speed. (Source: MDR)

the other external cache buses. (For small caches, the G4 will be offered with a 64-bit bus, halving its L2 bandwidth.)

Athlon, due to its 200-MHz system bus, and the G4, due to its 128-bit bus, offer the highest memory bandwidth (assuming it is matched by DRAM bandwidth). Motorola says the G4's bus is enhanced over the 750's to support frequencies well beyond 100 MHz, although it is not clear that Apple's system-logic chips will accommodate the higher speeds. The G4 will also be offered with a 64-bit bus for lower cost and for pin compatibility with the 750.

### Power Limits Markets

For desktop PC processors, power consumption, within reason, is not a big issue. Any vendor that has its sights set on notebook PCs, however, must pay close attention to power. As Figure 7 indicates, only Athlon and Mojave are hopelessly outside the power envelope of the portable market.

Intel intends to offer a mobile version of Coppermine. Based on the power dissipation of Katmai and the characteristics of its new 0.18-micron P858 process (see MPR 1/25/99, p. 22), we expect Coppermine to dissipate about 20 W (max, at 667 MHz and 1.5 V)—too much for use in notebooks. Using 1.1 V, however, which P858 is capable of, and a reduced frequency of 600 MHz, Coppermine can be brought within the 10-W thermal envelope Intel prefers for notebooks.

Although in its initial 2.8-V process WinChip 4 will not be suitable for notebook duty, this will be immediately remedied once the chip is rendered in a 0.25-micron process, as originally intended. In such a process, WinChip 4 should easily reach 500 MHz and dissipate less than 10 W (max).

Rise's mP6 II appears to be in the best shape to enter the portable x86 market. It will, however, be limited to the low end or the ultralight notebook market, as the chip's frequency limitations would be a nonstarter in traditional notebooks. Motorola's G4 is also well suited for notebooks,

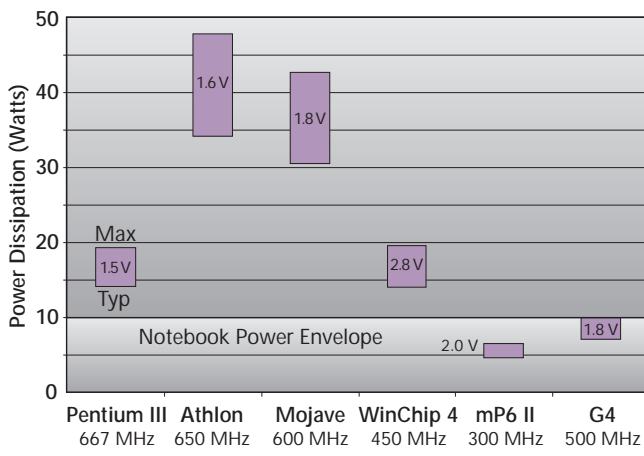


Figure 7. Since manufacturers are reluctant to quote power figures before chips are fully characterized and in production, we have estimated their power levels on the basis of previous processors from the respective companies plus the operating frequencies and voltages of the new processors. (Source: MDR)

even at speeds up to 500 MHz. PowerPC 750 processors have historically consumed much less power than their x86 counterparts, and Motorola's copper HIP5 process will help the G4 maintain that distinction, despite its higher frequencies and its transistor-intensive AltiVec execution units.

### The Price Must Be Right

In the x86 PC processor business, competitors that hope to make a profit must keep manufacturing costs low. Figure 8 shows how the various chips stack up on manufacturing costs. WinChip 4 and the mP6 II have a tough row to hoe: considering their frequency, performance, and lack of brand recognition, these chips will have to be priced at \$40–\$60, yet the parts have manufacturing costs that are not much lower. WinChip 4 will be in much better shape once in a 0.18-micron process, but Rise is stuck with a large die, having already pulled the 0.18-micron trigger. Mojave also has a dilemma, since its manufacturing cost will be about the same as Pentium III's.

Intel intends to make the situation even worse for these competitors. With the L2 integrated on chip, Intel has stated that it will offer Pentium III in a single-chip package, eliminating the bulky, costly SECC2 module. Flip-chip mounting to an organic PGA or BGA package is a likely scenario. This would completely eliminate the module costs and dramatically reduce package cost from that shown in Figure 8.

With Athlon, AMD is taking a different tack than other competitors. The company hopes its new processor will outperform Pentium III, justifying prices on a par with those of Intel's high-end parts. Even if the market goes along, however, AMD has no hope of matching Intel's margins, because of its much higher manufacturing cost.

To alleviate this problem, AMD must get Athlon onto its 0.18-micron process, which it hopes to do before the year is out. The 0.18-micron process will boost frequencies and will allow AMD to either lower the die cost or to add an on-

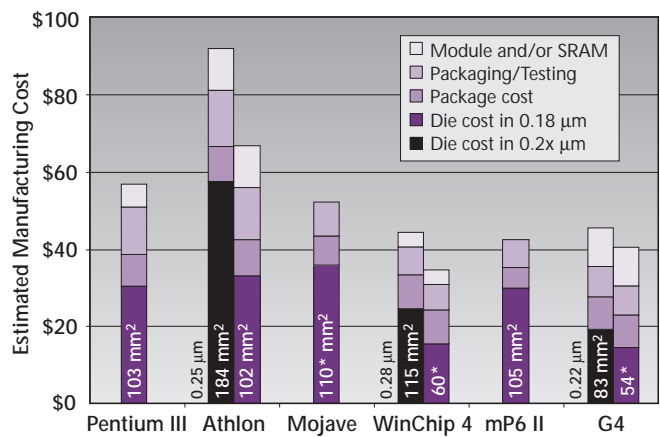


Figure 8. Both Athlon and WinChip 4 must quickly move to a 0.18-micron process to compete profitably with Intel in their respective segments. \*MDR estimate (Source: MDR Cost Model)

chip L2 cache, so it too can eliminate module and SRAM costs. Both options seem likely.

When we look at die costs only, Motorola's G4 wins hands down. The part owes its small die size primarily to its RISC architecture, which allows a simpler implementation than the x86 CISC architecture. The G4 will suffer, however, from a costly ceramic package, plus the cost of external cache SRAMs and some type of interposer or other module technology to mount them.

Moving the design to 0.18-micron HIP6 will lower die costs, but Motorola's larger problem is packaging costs. Given the G4's small die size, bringing the L2 cache on board seems like a good solution. This would eliminate the need for an interposer and allow migration to organic package substrates. Motorola will describe its second-generation G4 processor at Microprocessor Forum in October.

### Benchmarks Needed for Comparisons

The analysis presented here was performed in lieu of benchmarks, which are, unfortunately, not yet available for any of these chips. Although the data and charts provide some insight into the various microarchitectures, it would be inappropriate to jump from this analysis to any quantitative conclusions about the bottom-line performance of the chips.

In this limited analysis, we have focused on a few revealing factors, ignoring important effects such as TLB misses, excessive register dependencies, etc., which can substantially diminish performance, especially on x86 processors. These effects tend to have a disproportionately large impact on the more-parallel, higher-performance machines. Thus, our analysis probably exaggerates the differences between complex and simple processors. We expect, for example, that on benchmarks WinChip 4 will perform better relative to the other chips than is indicated by Figures 4, 5, and 6.

### AMD Seeks High Ground, Others Bottom Feed

On the basis of this analysis, however, we can roughly position the chips into the PC-market segments for which they are likely to be most suitable. The segments, defined by Intel, include performance-desktop (system price >\$2,000), mainstream-desktop (\$1,000 to \$2,000), value-desktop (\$500 to \$1,000), and mobile segments, as Figure 1 shows.

We expect Coppermine and Athlon alone to battle it out in the performance-desktop segment—Intel with a superior IC process, AMD with a superior microarchitecture. To deliver on the potential of its microarchitecture and to overtake Intel, AMD must quickly get to 0.18 micron, add an on-chip L2 cache, and enhance the design with SSE. The on-chip L2 cache will become increasingly important, as CPU clock rate increases will quickly bypass SRAMs, forcing Athlon to a one-third speed L2.

This analysis offers little hope that Cyrix, Centaur, or Rise will move very far up the food chain. Intel's aggressive IC processes and its occasional architecture tweaks (e.g., SSE) appear sufficient to keep the company out in front of

these competitors on performance, despite its aging P6 microarchitecture. If these competitors are able to keep their manufacturing costs under control, however, there is a several-million-unit market below the \$65 price point, which Intel appears happy to let others scavenge.

Mojave appears to have microarchitectural characteristics that might have allowed it to compete in the high-end mainstream segments. But National's CMOS-9 process would have precluded that, and now, with Via, it is unclear what fab process Mojave will use or how much jeopardy a fab change will put into its schedule. Therefore, realistically, we expect Mojave to compete only in the low-end mainstream and value segments, where Cyrix has traditionally been focused.

But these segments are a dangerous place to play, and Mojave's position is quite precarious. Assuming it can achieve 600 MHz by 2Q00, as promised, Mojave will be in place for only a short while before Coppermine comes crashing down upon it. Intel has made crystal clear its intentions to defend these segments with vicious price cuts. AMD also has its sights set on that market, and it will probably attack vigorously with 0.18-micron Athlons.

From a microarchitecture perspective, WinChip 4 seems best suited to the value-desktop segment. Centaur may have had its heart set on the high end of this segment, but, saddled with IBM's old 0.28-micron process, WinChip 4 seems destined for the low end of the value segment. If Centaur can find an owner and quickly move WinChip 4 to 0.18 micron and bring its frequency up to 600 MHz or more, the part might find its way into a few sockets in the upper end of the value segment. WinChip 4's primary advantage is that it can reach the lowest cost point of any of the x86 competitors, as it is clearly the simplest and most efficient of the designs.

Rise's mP6 II, whose bloated die size and low frequency make it an unlikely competitor in the desktop segments, may be able to carve out a niche in the low-end and ultralight mobile segments, based on its low power consumption. These segments are somewhat less cost-sensitive than the desktop segments, and there Rise may be able to get away without 3DNow or SSE. A 0.18-micron WinChip 4, however, may give Rise a run for its money in the mobile segment. Intel will claim the high-end mobile segments for itself with the mobile version of Coppermine, preventing either company from moving very far up the mobile food chain.

Although not a direct competitor in these segments, Motorola's RISC-based G4 delivers the most bang per square millimeter of silicon and per watt of power. Even though the part will never match Pentium III or Athlon in frequency, it can hold its own on general-purpose integer and floating-point applications, and it should significantly outperform both of those processors on multimedia applications. Its low power consumption also makes it well suited to notebooks. Thus, the G4 should keep Apple's Macintoshes competitive with Wintel machines through at least one more generation. 