Intel's Long-Awaited P55C Disclosed

First MMX Implementation Also Has Pipeline Tweaks, Larger Cache

by Michael Slater

For much of this year, the Pentium market has lived in the shadow of the much-anticipated P55C—the first processor to implement the MMX multimedia instruction-set extensions. The P55C is Intel's last pass at enhancing the Pentium line, which will still dominate PC shipments in 1997 but will give way to P6-family chips in 1998.

At the Microprocessor Forum, David Perlmutter, head of Intel's Israel Development Center where the P55C was designed, described the modifications made to the P54C. (Other accomplishments of the Israel center include the 8088, 8087, and 860XP processors, Ethernet controllers, and cache controllers.)

The P55C was designed to minimize the effort required to convert P54C systems. It is pin-compatible with the P54C, requiring only a different core supply voltage (2.8 V). In addition to the new MMX instructions (*see* **100301.PDF**), the P55C has twice as much on-chip cache memory as its prede-

Figure 1. Compared with its predecessor, the P55C adds two new function blocks (shown in dark purple): an MMX unit and a return stack buffer. Several other areas, shown in light purple, are enhanced from their functions in the P54C.

cessor and a host of smaller improvements: a slightly different pipeline structure, enhanced branch prediction, a new return stack buffer, and deeper write buffers. (The P55C is officially called Pentium Processor with MMX Technology—a name we find so catchy we're going to keep calling it the P55C.) Figure 1 shows the block diagram.

Intel has had silicon since late summer, but formal announcement of the chip won't occur until volume system shipments begin early next year. Many observers expected the chip to ship by mid-'96. The delay appears to have been in design, not debugging: Perlmutter said that only one erratum has been found since the first silicon was produced. This is due, in part, to a massive simulation effort, running hundreds of millions of cycles per week using the slack time of more than 3,000 workstations throughout the company.

Part of the delay may be due to the fact that Intel just didn't need the chip this year to keep its hold on the PC market, with both AMD and Cyrix making weak showings. The P55C could have boosted the slowing growth in the PC industry if it had shipped earlier, but Intel has done fine without it. By delaying the P55C announcement past the Christmas season, Intel keeps holiday purchasers focused on P54C and Pentium Pro systems. The delay also gives the software community more time to deliver applications that make use of MMX. A strategic benefit of the delay is that the lag before the P6 family gets an MMX version will be short, minimizing an awkward positioning situation.

The P55C initially will be available at 166- and 200-MHz clock rates and is likely to be priced at the high end of the Pentium market ($400–$500). As production ramps up, we expect Intel to lower P55C prices even more rapidly than P54C prices have dropped so the P55C (and therefore MMX) can sweep the earlier Pentiums aside by the end of 1997 or even sooner.

Revised Pipeline Adds MMX, Tweaks Speed

Integrating the MMX instructions into the P55C started with a complete redesign of the instruction decoder. In the P54C, there are no performance-critical instructions starting with the "0F" opcode used by all the MMX instructions; decoding these instructions in a single cycle required the decoder to be reworked. To provide more time for the more complex instruction-decoding task and ease some critical paths, the P55C's designers extended the front end of the pipeline by one stage, as Figure 2 shows. Instructions are read from the cache during the PF stage, while instruction parsing and prefix decoding is handled in the new F stage.

This change provides other benefits as well: it provides slightly more data-cache access time (another critical path in
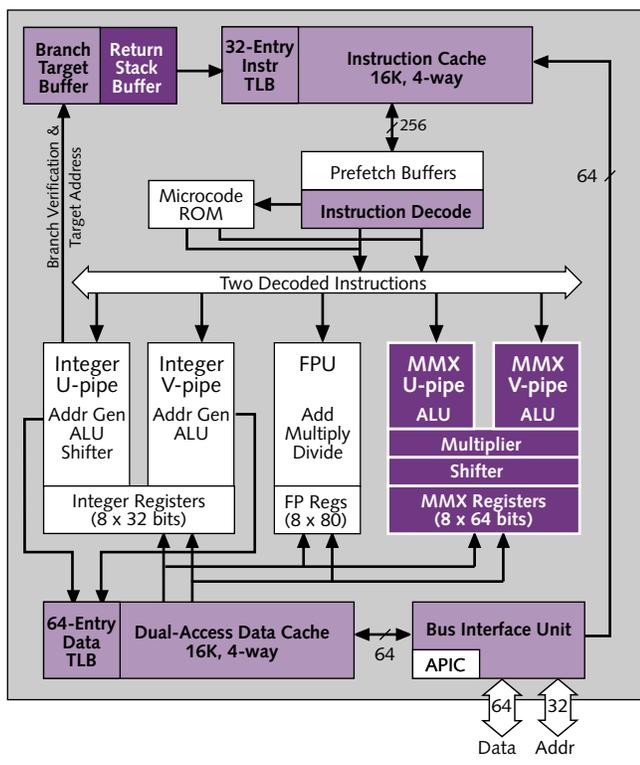
P54C

| | PF | D1 | D2 | E$_{1-3}$ | WB |

P55C

| | PF | F | D1 | D2 | E$_{1-3}$ | WB |

P55C MMX (non-multiply)

| | RD | E | WB |

P55C MMX Multiply-Add

| | RD | M1 | M2 | M3 | WB |

PF: Prefetch
F: Fetch (length decode)
D1: Decode and issue
D2: Address generation
E: Execute
WB: Write back
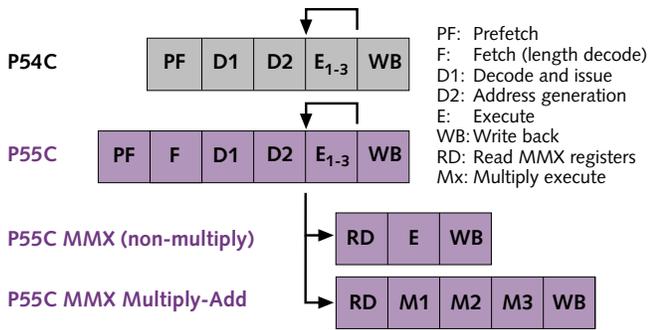RD: Read MMX registers
Mx: Multiply execute

Figure 2. The P55C adds one pipeline stage to P54C for standard instructions; for MMX instructions, the pipeline is extended by three more stages for multiply or multiply-add and one stage for others. Note that the E stage in the standard Pentium pipeline can be one, two, or three clocks long, depending on the instruction.

	P54C	P55C	Pentium Pro
L1 Cache (I/D)	8K/8K	16K/16K	8K/8K
MMX?	No	Yes	No
Peak Issue Rate	2 instr	2 instr	3 instr
Out-of-Order?	No	No	Yes
TLB Entries (I/D)	32/64	32/64	32/64
Branch Target Buffer	256-entry	256-entry	512-entry
Branch Algorithm	Single-level	Two-level	Two-level
Return Stack	None	4-entry	4-entry
Write Buffers	2-entry	4-entry	Not disclosed
Max Clock	200 MHz	200 MHz	200 MHz
Supply Voltage	3.3 V	2.8 V	3.3 V
Transistors	3.3 million	4.5 million	5.5 million
IC Process	0.35μ, 4M BiCMOS	0.28μ, 4M CMOS	0.35μ, 4M BiCMOS
Die Size	90 mm^2	140 mm^2	196 mm^2
Est. Mfg Cost*	$40	$60	$145†

Table 1. Intel P55C enhances the original Pentium design while boosting its performance with a larger cache, design tweaks, and MMX. †includes 256K L2 cache (Source: Intel, except *MDR)

P54C), so it should improve 200-MHz yield, and it allows non-MMX instructions to be paired for dual-issue in some circumstances that the P54C did not allow.

In the P54C, a tag bit is added to each byte as it is stored in the instruction cache to identify instruction boundaries. The P54C's decoder depends on this bit to feed the two instruction pipelines in a single cycle. The P55C's extra cycle allows instructions to be paired on the fly, eliminating the need for the cache predecode bits and allowing instructions to be paired even on an instruction-cache miss.

The longer pipeline has a cost, of course: one extra penalty cycle on a mispredicted branch. To mitigate this downside, the P55C has an enhanced branch target buffer that uses a two-level algorithm similar to the P6's (see **090405.PDF**). Also following the P6 design, a four-entry return stack buffer (RSB) was added to predict the target of subroutine returns.

MMX Issue Restrictions

Like the integer unit, the MMX unit has two independent pipelines. Instructions can be issued in pairs to the MMX unit with a few restrictions. Any two MMX ALU instructions (and, or, add, subtract) can be issued as a pair. Only one of a pair of MMX instructions can access memory or integer registers, perform a multiply operation, or perform a shift or pack/unpack operation. Unlike the integer pipelines, however, in which the shifter is fixed in the U-pipe, the MMX unit can switch the single shifter and multiplier units to either pipeline.

Intel chose not to integrate the MMX functions into the floating-point unit, even though that unit already has wide registers (which are logically the same as the MMX registers) as well as a multiplier and adder that potentially could be reused. Giving MMX an independent unit provided dual-issue of MMX instructions as well as allowing the power-hungry FP unit to be shut down during MMX operations. (Note that even though the MMX and FP registers are physically separate, they are logically the same; FP and MMX code cannot be intermixed.)

One integer and one MMX instruction can be issued together, subject to some restrictions (for pairing rules, see Intel's *MMX Developer's Guide,* Chapter 3, available from the Web address in the box on page 22). Floating-point instructions cannot be paired with integer or MMX instructions.

MMX instructions all have a single-cycle latency, except for multiply and multiply-add, which have a three-cycle latency but are fully pipelined for a single-cycle issue rate. There is a two-cycle load-use delay before an MMX register can be stored to an integer register or memory.

As in any Pentium processor, instructions cannot proceed out of order. If one of the integer or MMX pipelines stalls, the other pipeline stalls.

The P55C die is surprisingly large: at 140 mm^2, it is about 50% bigger than the P54C. The new pipeline stage, dual-issue MMX unit, and larger cache all add to the die size. Even so, the MDR Cost Model estimates the manufacturing cost to be only $60, about 50% greater than for the P54C. Future shrinks will reduce this cost as the part moves into the entry-level market in 1998.

Optimizing for Faster Core Clock

Intel doubled the cache size to reduce the performance lost from cache misses at high core clock speeds. In addition, the caches are four-way (instead of two-way) set-associative. Preliminary data from Intel shows the doubled caches cut the data-cache miss rate (on SPECint95) by 20–30% and the instruction-cache miss rate by 35–40%. The net effect, combined with the pipeline and branch-prediction enhancements, is a 10–20% increase on standard benchmarks, according to Intel. The benefit of the large cache will be greatest at the 200-MHz clock rate, due to the higher cost of cache misses at that speed.

Other minor changes were made in the write buffers and the TLBs. The write-buffer depth was doubled, from two entries to four. In addition, the protocol was changed to

better handle buses operating at a fraction of the core frequency; the original P5 design ran at the core rate, and although the P54C introduced a fractional-speed bus, the logic was not optimized for it.

The instruction and data TLBs in the P55C are fully associative, an enhancement from the P54C's four-way set-associative design. In addition, the separate eight-entry data TLB for handling 4M pages is eliminated; every TLB entry can now handle either 4K or 4M pages.

Intel Moving Back to CMOS

The P55C, built in a pure CMOS derivative of Intel's 0.35-micron BiCMOS process *(see 090905.PDF)*, marks the company's departure from the BiCMOS process it has used since the first Pentium. It has the same metal pitches as the 0.35-micron BiCMOS process, so its density is the same (and Intel still calls it a 0.35-micron process), but the drawn gate size is reduced to 0.28 microns to accelerate the transistors. The potential speed increase is countered, however, by the lower supply voltage and the lack of bipolar transistors. The net effect is that the P55's circuits are about the same speed as the P54C's.

Intel expects the P55C to consume slightly less power at a given clock speed than the P54C. On the positive side, it has a slightly lower core voltage than even Intel's low-voltage P54CS chips, power-hungry bipolar transistors have been eliminated, the gate length is smaller, and additional effort was made to trim power consumption (for example, the new instruction decoder is fully static). On the other hand, the larger caches, more complex pipeline, and new MMX unit increase the number of transistors that must be switched.

Intel did not disclose power-consumption figures but said that the 166-MHz P55C would fit within the power envelope of the low-voltage 150-MHz P54C (which consumes 3.8 W typical), making it suitable for high-end portables. The mobile versions of the P55C use a 2.5-V supply voltage rather than the 2.8 V of the desktop version, cutting

the power consumption by about 20% but limiting the clock speed. Whether a mobile P55C-200 will ever exist is unclear.

The P55C will be Intel's flagship mobile processor until the 0.25-micron P6-family processor, code-named Deschutes, is ready in late 1997. Even after Deschutes appears, the P55C will power lower-priced notebooks throughout 1998. A shrink of the P55C to the 0.25-micron process is possible, enabling the P55C to hit 200 or even 233 MHz within the current power envelope. Because of its radically different P6 bus interface, Deschutes will require new notebook motherboard designs, while the P55C will fit into existing designs. This advantage could extend the life of the P55C in the mobile market.

Positioning the '97 Lineup

By 2Q97, Intel will have a rich product portfolio: both P54C and P55C versions of Pentium, plus Pentium Pro and Klamath in the P6 family. Giving PC buyers so many choices will require careful positioning, however, especially since there are some awkward aspects of the 1H97 lineup.

MICHAEL MUSTACCHI

Intel's David Perlmutter describes the P55C microarchitecture at the Microprocessor Forum.

Today, Intel faces one clear positioning challenge: Pentium vs. Pentium Pro. It has tackled this by equating Pentium with Windows 95 and the consumer market while pitching Pentium Pro to go with Windows NT for business users. The P55C's increased performance, however, will reduce Pentium Pro's performance lead over Pentium, especially for Windows 95. For those users who are drawn to MMX, there may be some months after P55C debuts but before Klamath ships when you have to buy the older processor line to get MMX.

Another awkward aspect of the P55C rollout is that MMX is most attractive for entry-level systems (since it is a low-cost way to implement multimedia functions), but the initial MMX processors will carry premium prices. Intel probably couldn't service the demand for a $100 or even a $200 P55C in the first half of the year, so it must convince consumers to keep buying P54C systems in the lowest price brackets until it can ramp up P55C volume.

Intel must also convince consumers, who traditionally have focused on clock speed, that a 166-MHz P55C is comparable to a 200-MHz P54C, and that a 200-MHz P55C is even faster. Intel will use its iCOMP rating to drive this point home. Intel recently revised iCOMP *(see 100902.PDF)* to include multimedia code, which MMX will speed up. As a result, this rating will show a bigger benefit than the 10–20% expected on non-MMX applications.

By 1998, the clumsy transition to P55C and Klamath/Deschutes will be over. The P54C will be relegated to the dustbin, the P55C will hold down the low end, and Intel will drive as much of the market as possible to its next-generation P6-family processors. Ⓜ