Integration and Segmentation Intertwined

Intel Will Limit On-Chip L2 for Market, More Than Technical, Reasons

The debut of Mendocino this fall will mark the beginning of the end of off-chip cache memory in the PC market. This transition began with the introduction of Pentium Pro, which brought the L2 cache into the processor module. Now the L2 cache is well on its way to becoming an integral part of the processor chip. The rate at which this design approach moves throughout Intel's product line will be limited not so much by its technical merit as by its implications for fab capacity and its effects on Intel's market segmentation strategy.

With Mendocino, Intel is integrating a wimpy 128K L2 cache—one-fourth the size of the cache in a Pentium II module. The relatively small size is dictated in part by cost and volume concerns: a larger on-chip cache would, of course, result in a larger die, reducing total manufacturing capacity and raising manufacturing cost. As part of the Celeron line, Mendocino is a cost-focused chip, and it therefore made sense to use the smallest reasonable cache size.

The small cache size also helps maintain the differentiation between Pentium II and Celeron. Even with the small L2 cache, however, Intel must artificially limit Mendocino's speed to maintain the market segmentation. Since it is based on the same CPU core and process technology as Pentium II, Mendocino could probably run at 400 MHz at introduction. But to maintain clear positioning for Celeron and Pentium II, their clock speeds should not overlap. Mendocino initially will be available at 300 and 333 MHz; although Pentium II will persist for a short time at these speeds, we expect Intel to shift almost all of its Pentium II business to 350 MHz and up. This will leave Celeron's use of a 66-MHz bus as an additional differentiator, since Pentium II processors at 350 MHz and higher use a 100-MHz bus.

To keep the low-price Celeron processors from eating into the higher-profit Pentium II market, Intel can't increase the Celeron speeds until it has a new part (or new speeds) to bring in at the high end of the line, allowing the former high-end Pentium II to become the midrange and the entry-level Celeron to move up to the performance position formerly occupied by the slower Pentium IIs (which then become obsolete). The introduction of Katmai in early 1999 will give Intel the tool it needs to make this transition.

When Intel ramps up its 0.18-micron process technology in the second half of 1999, the opportunity to increase the size of the on-chip L2 will be compelling. Even with an on-chip L2 of only 256K (expected in a part code-named Dixon), performance on most applications will probably be better than with a half-speed 512K cache off chip, and the manufacturing cost will be lower. In workstation and server applications, multimegabyte off-chip caches may be appropriate, but the performance gain they offer for typical PC applications does not justify their cost. Thus, starting with the 0.18-micron generation, processors with on-chip cache probably cannot be limited to the entry-level Celeron line.

This leaves the question of what will differentiate the Celeron line. Just selling the same chip under a different brand name for the slower clock speeds doesn't make much sense (but is not beyond possibility in this age of consumer marketing of microprocessors). Intel might integrate smaller caches in the Celeron parts. It also might leave the Katmai New Instructions out of Celeron for a time.

The next step in the differentiation strategy may be to integrate the north bridge and graphics functions onto the processor. With 0.18-micron technology, doing so with a Katmai core should be feasible; in 0.13-micron technology, which should begin production in 2001, the cost and performance advantages should be overwhelming. Most of the entry-level market will not care about the latest advances in graphics, and a single Direct RDRAM interface integrated on the processor will provide a very effective memory interface for such systems, with a reasonable number of pins.

In 0.10-micron technology, a single-chip PC will be very attractive. Available transistor counts will grow faster than the demand for those transistors in the CPU and cache, at least in entry-level PCs, making integration inevitable. Just as the MMU, FPU, L1 cache, and now the L2 cache have moved on chip, so will the system logic. The only thing that could slow this trend would be new mainstream applications that demand dramatically more CPU and graphics speed.

The initial Celeron may seem like a foolish product, but, in time, the segmentation strategy will make more sense. Intel would not have invested in launching a new brand name if it did not believe it had a long-term future. Today, Celeron is distinguished by the lack of L2 cache. In the near term, it will be distinguished by a smaller L2 cache and a slower system bus. Next, as an interim step, Intel will position the Whitney chip set, which combines the graphics and north bridge into one chip, as a companion device for Celeron processors. And in the long run, Celeron processors are likely to be the first Intel chips since the 486SL to integrate peripheral functions with the processor. M

See www.MDRonline.com/slater/segments *for more on this subject. I welcome your feedback at* mslater@mdr.zd.com.