Shift to On-Chip Cache Pays Off

Intel's Slot 1 and Slot 2 Will Give Way to Sockets by 2000

Intel's decision to release its Mendocino processor without a module in 1Q99 is just the tip of the iceberg. By the end of next year, we expect Intel to be shipping moduleless processors into all of its market segments, and by the end of 2000, virtually all of its chips will plug into sockets instead of slots. This trend will be enabled by a shift to on-die level-two (L2) cache, which makes today's module structure superfluous.

The initial purpose of the Slot 1 module was to hold the external L2 cache chips required by the Klamath and Deschutes CPUs. Mendocino (see MPR 8/24/98, p. 1) doesn't need external cache chips, as it is Intel's first processor to incorporate the entire cache subsystem. To maintain compatibility with these earlier processors, Mendocino is currently shipping in a Slot 1 module, despite the fact that, other than the CPU, the module contains no active components and is nearly empty.

Mendocino will discard this vestigial module early next year, instead appearing in a 370-pin plastic PGA that plugs into a standard socket. Intel will continue to offer the Slot 1 version as well, but we expect PC makers to convert to the so-called Socket 370 over time, as it is smaller and less expensive than Slot 1. By the end of 1999, the Celeron segment will be largely converted to Socket 370, and future Celeron parts are likely to appear only in a socketable form.

Intel will benefit from this transition as well. According to the MDR Cost Model, the PPGA version of Mendocino will cost about $10 less to manufacture than the current module. According to our shipment forecast for 1999, that decrease could save Intel up to $300 million next year.

Not one to let a good idea go to waste, Intel will eliminate modules from its other product lines as well. On-die L2 cache is very attractive in the mobile market because it reduces power consumption and physical size as well as cost. Sources indicate that in early 1999 Intel will deploy a mobile product code-named Dixon that has 256K of L2 cache on the chip. Given its power and size advantages, Dixon is likely to displace the current Mobile Deschutes fairly quickly, eliminating the comparatively bulky minicartridge.

Process technology is holding back the other segments. In the current 0.25-micron technology, even Mendocino's piddling 128K cache takes up about 35 mm^2 of die area, adding $10 to the manufacturing cost of the chip. Dixon's cache is likely to double that cost overhead, eliminating any cost advantage over a module.

In a 0.18-micron process, however, a 256K cache will fit into that same 35 mm^2, making it less expensive to integrate that amount of cache than to add an external cache. For this reason, we expect Intel's Coppermine processor, a 0.18-micron version of Katmai, to incorporate an on-die 256K cache. With on-die cache, Coppermine could plug into the same Socket 370 as Mendocino, although Intel will probably also offer a Slot 1 version as a transition vehicle. We expect Coppermine to appear in 2H99, and by mid-2000, it could largely eliminate Slot 1 from the PC market.

Even the Slot 2 segment could move away from modules. Intel says that future high-end processors Cascades, Foster, and McKinley will include on-die L2 caches as large as 2M (see MPR 10/26/98, p. 16). Like Mendocino, Cascades is likely to appear in both slot and socket versions, but Foster will probably use only a socket. Because servers and workstations are less sensitive to the cost and size penalties of the module, the transition to sockets will take longer in these segments than in the PC segments.

The 0.18-micron process is required in this segment as well: regardless of cost, the 0.25-micron process is simply incapable of building a processor with 2M of on-die cache, which is required to match the current Xeon line. Again, the cost savings of on-die cache are considerable: we estimate a 2M Cascades, even at a projected die size of 375 mm^2, will cost 50% less to build than today's 2M Xeon module, given the high cost of the Xeon's four custom cache chips.

Other vendors are also putting large caches onto their processors. AMD's next K6, known as Sharptooth, will include a 256K L2 cache. IDT will instead use 128K of primary cache (see MPR 12/7/98, p. 18). These chips will plug into Socket 7 but, like the Socket 370 parts, will not need external L2 cache, thus reducing system costs. The big exception to this trend will be AMD's K7, which will introduce the Slot A module just as Intel is shifting its products to Socket 370. Once the K7 moves to 0.18-micron technology, it too is likely to adopt on-die L2 cache, turning Slot A into Socket A.

The big losers in this transition are SRAM vendors, which won't have the PC market to sell into much longer. But shifting SRAM production to the processor vendors will require them to add fab capacity. Building this capacity will take time, slowing the transition to on-die cache. But build it they will. In two years, modules will be a fading memory.　Ⓜ

Linley Gwennap