



This document provides an overview of the P6 product, and P6 marketing programs.



The 5.5 million transistors in the P6 CPU deliver performance estimated at 200 SPECInt'92 (measured on a non-production system). And like all previous generation Intel processors, the P6 is fully binary compatible with existing PC software.

As discussed in other parts of this CD, the P6 was developed with a system-level design approach. The processor, L2 cache, and chipset were all designed and tuned to work together. This means that even the very first P6-based systems will be designed to take advantage of P6 performance. One example is that the P6's L2 cache works at the same speed as the processor itself, which reduces waitstates that would otherwise reduce system performance.

Production of the P6 will start on Intel's stable 0.6 micron process technology -- the same process being used for current Pentium® processors.



The first P6 has been sampling since January '95. It features a tightly coupled 256K full speed L2 cache in the same package as the processor. Also included are 8K L1 code and data caches. All of the P6 caches are non-blocking. This solution delivers caching performance equivalent to a typical 2MB external, blocking L2 cache.

This P6, again - built on a .6 micron process, will operate at 133MHz. The P6 will be moved to a .35 micron process in the future, enabling even higher performance (with planned frequencies above 200MHz).



The basic concept of dynamic execution entails allowing the processor to continue to usefully execute instructions while data for previous instructions is being prepared for their use. The three key concepts are multiple branch prediction, data-flow analysis, and speculative execution.

The non-blocking caches allow outstanding cache misses to be propogated out to the system bus for completion but still allow subsequent instructions to gain access to their data in the caches. The transaction oriented bus continues the non-blocking aspects of the architecture out into the system.

Out of order execution enables the processor to view a large pool of available instructions and data-flow analysis allows it to select the ones that are ready from a data perspective and dispatch them to execution units.

Speculative execution and multiple branch prediction is the way in which the "large pool" of instructions is filled. The processor will find a path through the code that it has a high confidence level will be correctly followed and it prefetches and executes as much of it as it can.

Register renaming is a way of providing the Intel Architecture with a large pool of temporary registers in a way that is transparent to application software.



When P6 systems appear in the second half of 1995, you'll see a broad range of performance and capabilities. Starting with desktop PCs, you'll see premium desktops with robust features such as advanced graphics, audio, 1GB hard drives, etc. Up from there, you'll see a number of workstation class systems with up to 2 P6 processors. These dual processing machines will be superb for WinNT and/or Unix.

A key focus for P6 has been to enable 4-way MP systems, so you'll see numerous 4-way P6 servers at aggressive prices. At the very high end, you'll see clusters of 4-way P6 systems forming super-server and mainframe class solutions.

As you can see, the P6 is designed as a scalable architecture. This opens a broad range of new software opportunities on the Intel Architecture: from 3D, multimedia, and desktop publishing apps that take advantage of DP, to client/server solutions for corporations.

<sup>\*</sup>other brands and names are the property of their respective owners.



In 1995 and 1996 the P6 will be targeted at the performance desktop and volume server market segments.

On the desktop, the P6 will first appear in premium machines. This market segment is about 10% of the total market segment for desktop PCs, which amounted to 33Mu in 1993 and has grown considerably since then. (This includes all Intel architecture compatible and non-Intel Architecture compatible desktops. Notebooks and servers are excluded. Data is supplied by Dataquest.)

In servers, the P6 will be in a broad range of machines. From \$4K base systems configured much like the desktop PCs, to \$100K machines with gigabytes of RAM and 10's of disk drives. This part of the market segment comprises almost half of the total server market segment according to Dataquest, and was 560Ku strong in 1993. We will also see some servers priced above this range, but in terms of units their numbers are small.



The P6 will enable unprecedented performance in desktop PCs. The diagram above gives you a closer look at a P6-based performance desktop and what features will typically be included.

There is a class of customers who want *more* than the volume desktop can deliver, and are willing to pay for it. These power users are continually looking for the most performance they can get to run their high end applications. Examples of uses include clients on an enterprise network, full-featured Multimedia Authoring systems, demanding databases, desktop publishing, CAD, animation/rendering, and software development.

Beyond these most demanding users, the P6's performance also opens up new capabilities that will be used by a broader audience of users. These include software-only video conferencing, continuous real-time speech recognition, broader use of 3D and visualization tools, and others.



This is an example of the kind of P6 server configuration that will be common. P6 will bring "PC hardware economics" into the volume server market segment. We use the term Standard High Volume Server to describe this market segment. Although the configuration show here will be "typical," you can expect many different system implementations. Note for instance that this low-to-midrange solution can be clustered to create a high-end "Enterprise" solution. All of the required "high-end" features are enabled into this Standard High Volume Server market segment.

Among the P6's server-enabling technologies, the P6 makes more extensive use of ECC than the Pentium® processor, providing ECC on memory/cache and ECC or parity on external buses.



The P6 core communicates with its L2 cache on a dedicated, private bus. This results in all CPU-to-cache traffic being contained within the P6 package. The external bus is a cache-to-memory bus and is only used to service L2 cache misses and IO requests. The P6 external bus is a standardized, SMP system bus (standardized since it is implemented in silicon). Remember that this external bus is also a transaction bus so there are no wasted cycles between the request and response phases.

The P6 bus interface unit snoops all memory transactions on the P6 external bus using the MESI protocol to keep its internal caches coherent in a multi-processing (MP) environment. The P6 supports cache-based semaphores which will dramatically reduce bus traffic when synchronizing multiple processors.

Even with the most stressful of benchmarks that we have simulated, the P6 external bus is less than 25% utilized. This will allow 4 P6's to be simply connected together to form an MP system. No external glue logic is required, the 4 CPUs are all connected in parallel and they operate in parallel using a standard SMP paradigm.

So the P6 integrates all the system-level components into a "single unit of MP". By integrating the cache controller, cache interface, cache SRAM, and APIC, we have a much simpler Multiprocessing design. A simpler, lower cost design will ensure that more MP machines are introduced to the market at lower prices, driving up the demand for MP applications.

What does this mean for **desktops**? If the system doesn't use 4 CPUs, then the 'spare' bandwidth can be used for IO devices. With 2 CPUs, for example, we still have 50% = 250MB/s of bandwidth which is two maxed out PCI buses. That's four 30 frame/sec full screen videos.

