Week 4: CPU architecture-3

In order to improve computational efficiency, there are a number of design features to modern processors that maximise their performance.

Instructions per clock (vs. clock speed)

Each type of processor is designed and built differently, with their own set of priorities and concessions: one of these is how much work they will perform per "clock tick" (the speed of the "clock ticks" is what is measured by the Megahertz/Gigahertz rating.)

The amount of work performed per clock tick is referred to as a processor's **Instructions Per Clock (http://en.wikipedia.org/wiki/Instructions_Per_Cycle)** (IPC). In recent computing history, the most famous architectural showdown between these two design concepts have been Intel's **Pentium 4 (http://en.wikipedia.org/wiki/Pentium_4)**, used in PC systems in the early- to mid-2000s; and Motorola's PPC7400, popularly known as the "**PowerPC G4 (http://en.wikipedia.org/wiki/PowerPC_G4)**" in Apple Macintosh systems from the same time frame.

This difference in design was made public by Apple, as the "megahertz myth". The basis of this myth is that the clock tick speed alone isn't a reliable indicator of performance -- it all depends on how much work actually gets done per clock tick.

The Pentium 4 processor sacrificed IPC performance for raw clock speed. At each clock tick, the processor wasn't able to compute much; however they ran so fast (at times, twice the speed of competitors) that in the end, their performance was equivalent to other contemporary offerings. This is an example of a processor with a low IPC, but high clock speed.

> This is why the Pentium 4's pipeline was designed to be so deep.
>
> The larger each stage of a pipeline is, the longer it takes to do its job. So, with its very long pipeline of small micro-ops, it was relatively easy to make it go fast.

Motorola's 7400 processor was designed with the opposite approach: the clock was to run slowly, but more work would be performed per clock tick. This processor had a high IPC, but low clock speed.

Each approach has its advantages and disadvantages, which eventually led to problems for both processors:

- faster clock speeds consume more energy. As the clock speeds increased, the Pentium 4 started to consume so much energy in such a little space that the excess heat generated could not be effectively dispersed.
- a higher IPC (that is, a processor doing more work each clock tick) means that the clock speed can't be increased quite as readily because of all the inter-dependencies and timing constraints inherent in a complex 'chunk' of transistor logic. The G4 CPU soon reached a design limit where the clock speed simply couldn't be increased.

Instruction level parallelism

Instruction level parallelism (http://en.wikipedia.org/wiki/Central_processing_unit#Instruction_level_parallelism%23Instruction_level_parallelism) is all about trying to get the CPU to do multiple things in parallel. The common method for achieving this is with instruction pipelining. Processors that exploit this technique are known as **superscalar (http://en.wikipedia.org/wiki/Superscalar)** architectures.