Chapter 2 – More Computer Evolution and Performance

Designing for Performance (Section 2.2) on pp. 38-45

The major breakthroughs came in the third generation. After this generation, there is no real agreement what defines another generation. As was mentioned earlier, the third generation was characterized by SSI & MSI and the production of the integrated circuit. The first major applications of the IC were: (1) construction of the processor (control unit and ALU) and (2) construction of memory.

Semiconductor Memory

In the 1950s and 1960s, the type of memory produced was called "core" memory because the memory was constructed of rings that could be magnetized in one of two directions. For its time, this memory was fast. Access time was in the neighborhood of a millionth of a second (1 microsecond). This memory tended to be expensive and used destructive readout which means that the data once read, erased the data stored at this location and therefore, it needed to be restored.

In 1970 the first semiconductor memory as we know it today was made. Access times were in the 70 nanosecond range and chip densities consisted of 256 bits per chip. Since 1970, memory chip densities have gone through several generations: 1K, 4K, 16K, 64K, 256K, 1M, 4M, 16M, ..., transistors per chip.

Moore's Law – Gordon Moore, cofounder of Intel, predicted that the number of transistors that could be put on a chip would double every year based on his initial observation of this phenomenon. Since 1970, this doubling pace has happened every 18 months and has remained consistent ever since.

The consequences of Moore's Law are:

1. The cost of a chip has remained virtually unchanged during this rapid growth.
2. The electrical path has been shortened because the logic elements can be placed closer together thereby increasing the operating speed dramatically.
3. Computers are becoming smaller making them more convenient to place in different environments.
4. There is a reduction in power & cooling requirements.
5. Interconnections on a chip are more reliable that solder connections between chips.

You are responsible for reading the history of the IBM System/360, Dec PDP-8, and Intel processors on pp. 32-38.

Intro to Performance Issues

Every year the cost of computers goes down while the performance goes up. How does this happen?

1) Microprocessors get faster through:

a) Branch Prediction – the processors look ahead and predict which instructions will be executed next.

e.g.
```
if (x < y)
{
  printf ("%d%d", x, y);
}
else
{
```

```
    printf ("%d%d", y, x);
  }
```

b) Data Flow Analysis – the processor schedules instructions in an optimized manner based on which instructions are dependent on which other instructions.

P#1: Give a C code example that demonstrates data flow analysis and how the processor can take advantage of scheduling.

c) Speculative Execution – instructions are executed ahead of their actual appearance in the program.

2) Performance Balance – different parts of the computer operate at different speeds. It is important to exploit as much concurrency as possible and keep the processor as busy as possible.

The following diagram shows processor speeds versus data transfer speeds between CPU & memory over time.

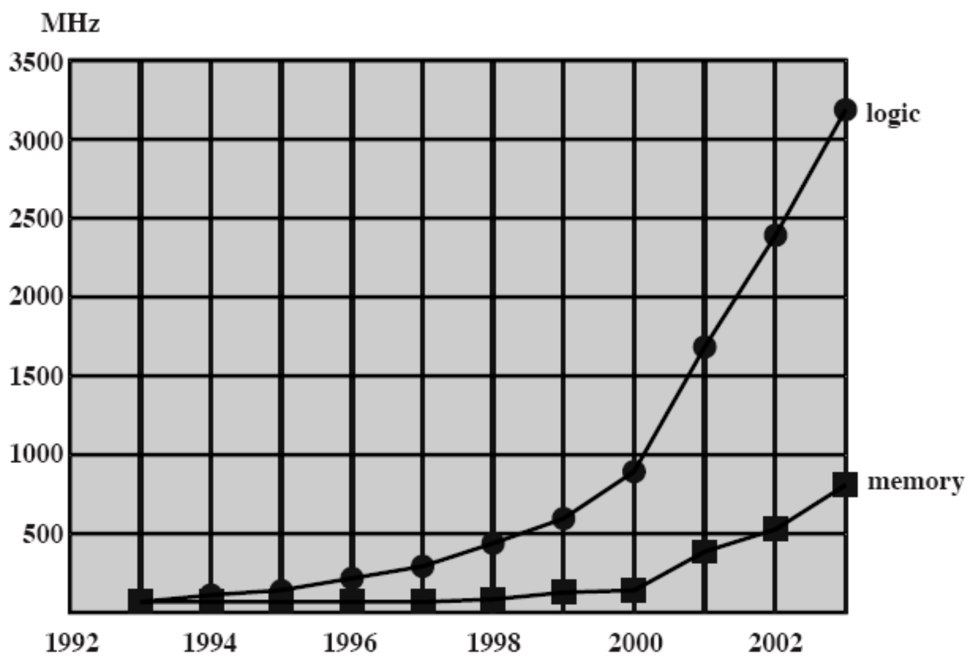

Figure 2.10 Logic and Memory Performance Gap [BORK03]

Q#1: How might you reduce the gap between logic and memory in the above figure?

3) Improvements in Chip Organization and Architecture

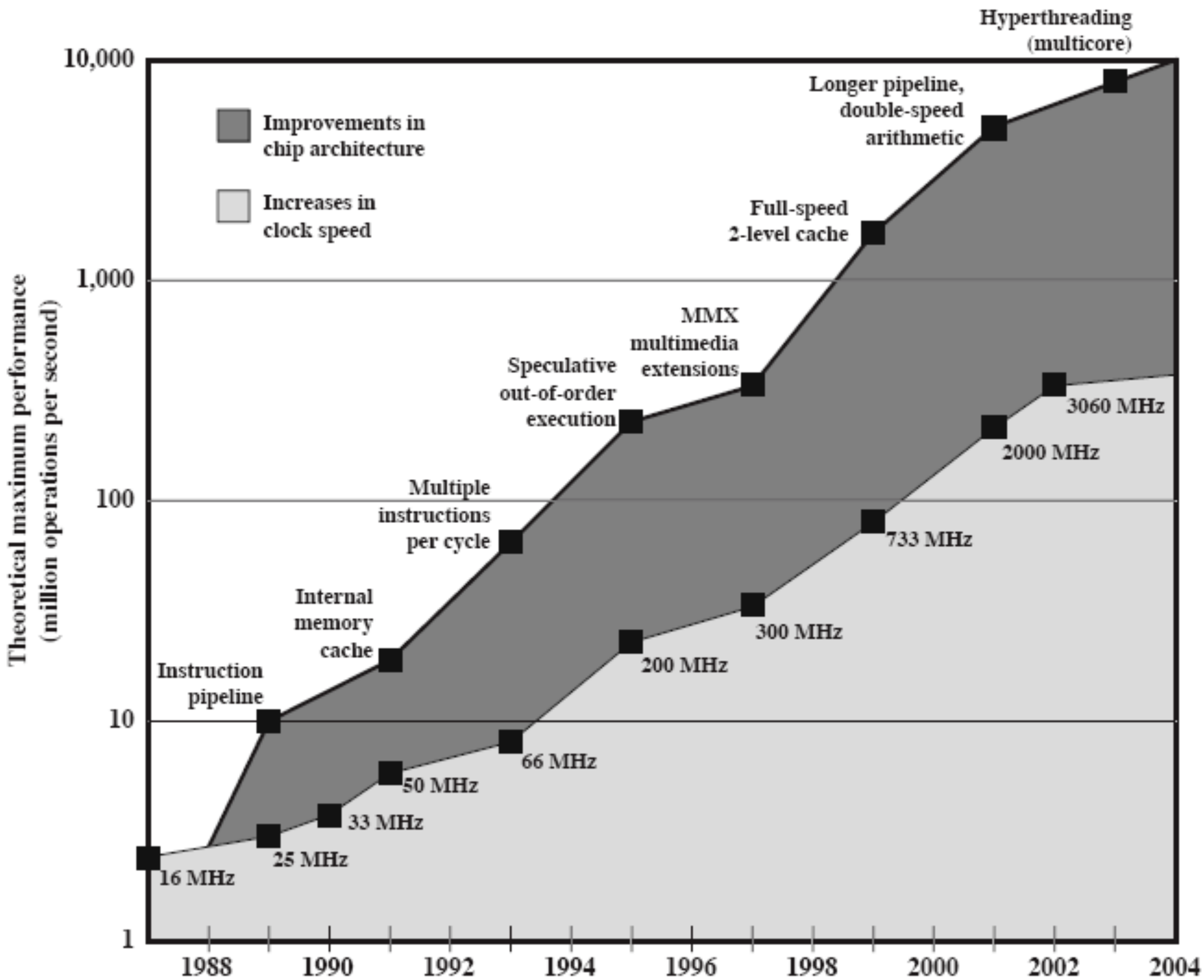

Figure 2.12 Intel Microprocessor Performance [GIBB04]

Note: We will spend a good deal of this course discussing the specifics of these three performance issues.

Pentium and PowerPC Evolution (Section 2.3) on pp. 45-48

You are responsible for reading Pentium and PowerPC Evolution on pp. 45-48.