Computer Architecture Assignment 1

Neale Ratzlaff

30 September 2015

1 Difference between architecture and organization

Organization is what most people think of when they think of architecture, when organization is mostly defined by the ISA. Organization is the physical layout of the hardware. The layout of the memory stack, how the ICs are alligned onboard are all part of the organization of the computer. This is important as data transfer speeds, power consumption, and system functionality are all affected by the organization.

Architecture is essentially the computer's instruction set. The set of rules that define how the processor will behave, which in turn necessitates organizational features. The architecture may also include the microarchitecture of the processor, or how the processor will implement each instruction in the ISA (a black box to users).

2 Endianness

Endianness is the concept of ordering with respect to bits in the system There are two types of endianness in modern computers, big-endian and little-endian In a big-endian system, the smallest memory address stores the most significant byte of a memory word Little-endian systems store the least significant byte in the largest memory address. Intel's x86 architecture uses little-endian representations. Little-endian systems are also common with microprocessors due to Intel's influence. Big-endian systems are most commonly found in computer networking applications. There is also another type called a mixed-endian system. A mixed system will have a different endianness for 16bit words vs. 32 bit words. Bi-endian processors can operate in little-endian or big-endian mode.

3 Floating point format (IEEE)

Single precision format is known as binary32. Binary32 consists of:

- 1 bit sign bit
- 8 bit exponent bit
- 24 bit significand component (23 explicitly stored)

Double precision format is called binary64:

- 1 bit sign bit
- 11 bit exponent bit
- 24 53 significand component (52 explicitly stored)

4 Memory heirarchy

The memory heirarchy of the computer refers to the arrangement of the memory locations of the system. The smallest and fastest memory cache is the L1 cache. L1 is further split into two levels of instruction and data cache, each 128 KiB. The instruction cache contains a block of opcode that has been fetched from memory. By temporal locality it can be assumed that code that instructions that have been executed recently are more likely to be used again, so having rencent instructions in the L1 cache allows for increased loop execution. The L1 data cache operates on the same principal, only it stores program data instead of instructions. The L1 data transfer speed reaches 700 GiB/s.

The next cache in the heirarchy is the L2 cache. The L2 is a unified data and instruction cache 1 MiB in size. L2 is a slower and farther away cache than L1, best access speed is 200 GiB/s. The L2 cache normally feeds data into the L1 cache.

The last level found in generally all systems is the L3 cache. This is the largest cache still at 6MiB. L3 is generally a shared cache between all the cores of the processor. L1 and L2 usually are dedicated to each core to improve performance. While L3 is a large shared space used as the big-brother backup to the smaller L1 and L2 caches. Access speed of the L3 cache is around 100 GB/s.

Some processors have an L4 cache that is the largest at 128 MiB, and acts as just one more level of cache memory. Beyond this is only main memory.

Flip has an L1 instruction/data, and an L2 cache.

5 SIMD Instruction Levels

Flip supports the following SIMD instructions :

- Streaming SIMD extensions 3
- Streaming SIMD extensions 4.1
- Streaming SIMD extensions 4.2
- Streaming SIMD extensions 4.1
- Streaming supplemental SIMD extensions 3