CMP SCI 635: Modern Computer Architecture

# Hitting the Memory Wall: Implications of the Obvious

Name: Kunjal Panchal

Student ID: 32126469

Date: 24th Sept, 2019

Paper: 6 – Memory [Wulf95]

Strengths:

1. The idea of a widening gap between memory and CPU speeds is something to think about. As a consequence, some programs will not scale with CPU speed improvements (memory-bound programs); some other programs may become memory-bound as CPUs get faster; one could say that these programs hit the memory wall.
2. While the processor heavy applications like multimedia servers won't be affected drastically by this CPU vs Memory speed exponents; the commercial applications where most of the time is spent in fetching and storing data will be able to make good use of this phenomena, where most of the times, CPU is idle because it is waiting for data.

Weaknesses:

1. The claim that compulsory misses (fixed miss probability) are inevitable: We can reduce the number of compulsory misses by, e.g., making the cache lines longer. Cache miss rates in general are very dependent on the application, and this holds even more for compulsory misses. So, the fixed miss probability of at least 0.2% doesn't have a firm ground.
2. On current high-performance machines the time spent in compulsory cache misses is limited to a few seconds if the process stays in physical RAM (and once you get into paging, compulsory cache misses don't have that much of a significance). Compulsory cache misses occur once, at system start-up time, if at all ("*several architectures can clear architectural memory without touching DRAM; subblock validate bits are another alternative to deal with writes*"). New processes just reuse cache lines for physical memory that is already in the cache, and don't suffer any compulsory misses.

Questions/Assertions:

1. Can a modified compiler (some totally new concept driving it) help in cache misses? Can we design applications in such a way that we don't have to worry about cache misses at all?
2. Does cache having multiple levels mitigate the effect of this Von Neumann Bottleneck?