Three Definitions

Introduction
In this assignment I will be giving a parenthetical, sentence, and expanded definition of a relatively complex term in computer science. The term that I have chosen to define is 'cache memory'. The goal of this assignment is to understand how audience and purpose indicate the need for definitions. It also aims to differentiate between the levels of detail in definitions in different situations, and how this is useful in technical writing.

Term
Cache memory

Parenthetical definition
Cache memory (a type of computer memory)

Sentence definition
Cache memory is smaller, faster, short-term memory in computers used to speed up the computer.

Expanded definition
Operating Principle
Cache memory works by using a smaller but faster memory storage in addition to a larger and slower memory. When the processor needs to search for some data in memory, it first looks to the cache to see if it is there. If it is not there, it looks to the next slower level of memory and this is called a memory miss (Jacob, 2008). There are many varying rules as to which data should be put in the cache (the most recent data vs the most used data) since it typically has a small amount space. There is no one 'optimal' way to decide what data should be put in cache memory without telling the future, but the best strategies seen so far use a combination of strategies. Modern computers are programmed to look for patterns in data usage, such as if you always use groups of data together, and these are put in the cache together (Kidd, 2015). The idea is to have the data that you need in the highest/fastest cache as much of the time as possible since this is faster than always looking in main memory. There can also be multiple levels of caches as seen in *Figure 1*, with increasing space and slowness as you get down to main memory.

History
Cache memory was first used in the 80's when the gap between the speed of the processor and the speed of using main memory grew to a noticeable size (Jacob, 2008). As technology gradually improved, the speed of processors increased, but the speed of main memory could not increase at the same rate as the processors because it was too expensive (Wu et al., 2018). This slowed down the entire system since the processor was being held up by the slow lookup/storage time of the memory (Wu et al., 2018). Cache memory was invented as an intermediate memory between the processor and main memory. It allowed an overall faster search time but still being affordable. There was not one person who invented cache memory, since it was a gradual shift in memory getting faster.

Visual
The following visual shows a memory hierarchy where cache memory is shown to be smaller and faster than main memory. The register file is in the processor, and each level below represents a level of memory. 'Cache memory' refers to all memory between the register files in the processor, and main memory. In this example, there are two cache levels (L1 and L2), and different computer designs can have different amounts of levels. The processor looks for data in the fastest but smallest memory level (L1) and if it does not find it, looks in the next level (L2) until it eventually gets to main memory.
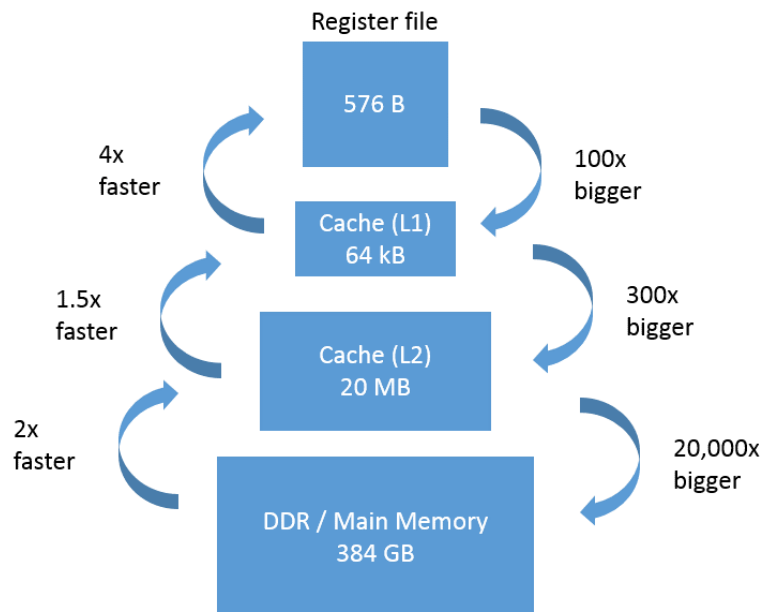
Figure 1. Diagram of Memory Hierarchy for Computers (Kidd, 2015)

Comparison and contrast
Cache memory and main memory are both forms of computer memory. When compared to each other cache memory is faster and is also more expensive (Przybylski, 1990). This differs from main memory, which is slower, but also cheaper. A computer with all cache memory would be too expensive to make, and a computer with all main memory would be too slow for users to use (Przybylski, 1990). A compromise is implemented where the benefits of both cache and main memory are utilized.

References

Jacob, B., Ng, S., & Wang, D. (2008). *Memory systems: cache, DRAM, disk.* Burlington, MA: Morgan Kaufmann.

Kidd, T., (2015). Advanced Computer Concepts for the (Not So) Common Chef: Memory Hierarchy: Of Registers, Cache & Memory. *Intel Software Developer Zone.* Retrieved from: https://software.intel.com/en-us/blogs/2015/06/11/advanced-computer-concepts -for-the-not-so-common-chef-memory-hierarchy-of-registers

Przybylski, S.A. (1990). *Cache and memory hierarchy design: a performance-directed approach.* Burlington, MA: Morgan Kaufmann.

Wu, W., Li, M., Wang, K., Huang, H., & Chen, E. (2018). Speed scaling problems with memory/cache consideration. *Journal of Scheduling, 21*(6), 633-646. doi:10.1007/s10951-018-0565-1.