NC State University
Department of Electrical and Computer Engineering
ECE 463/521: SPRING 2017
Project #1: Cache Hierarchy

by
Michael Patel

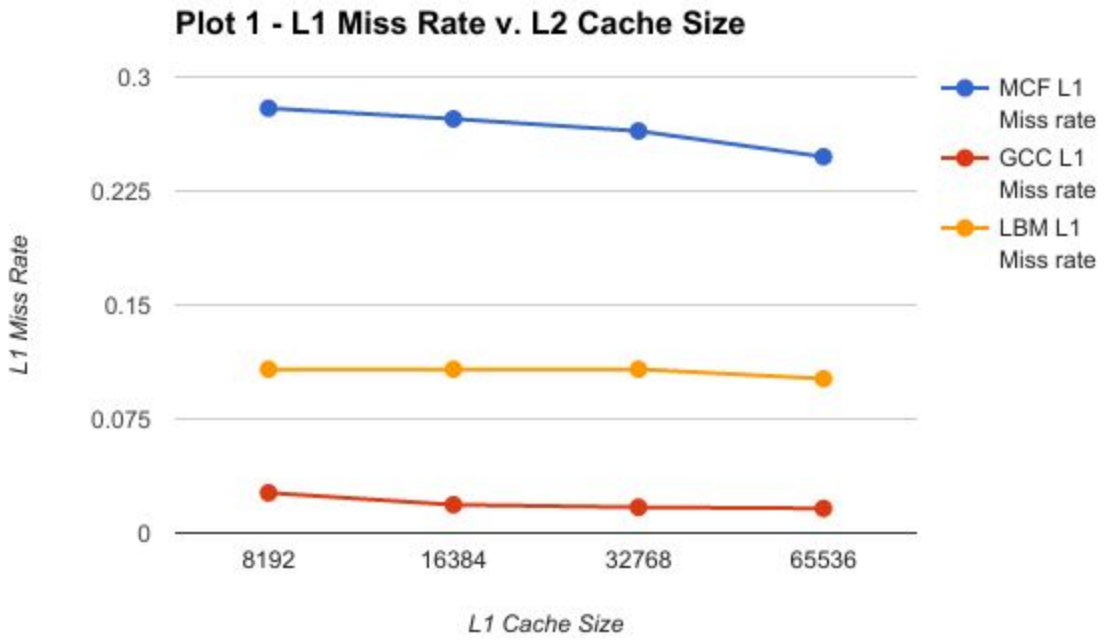
NCSU Honor Pledge: "I have neither given nor received unauthorized aid on this test or assignment."

Student's electronic signature: ____Michael Patel_____
(sign by typing your name)

Course number: ___521_____
(463 or 521?)

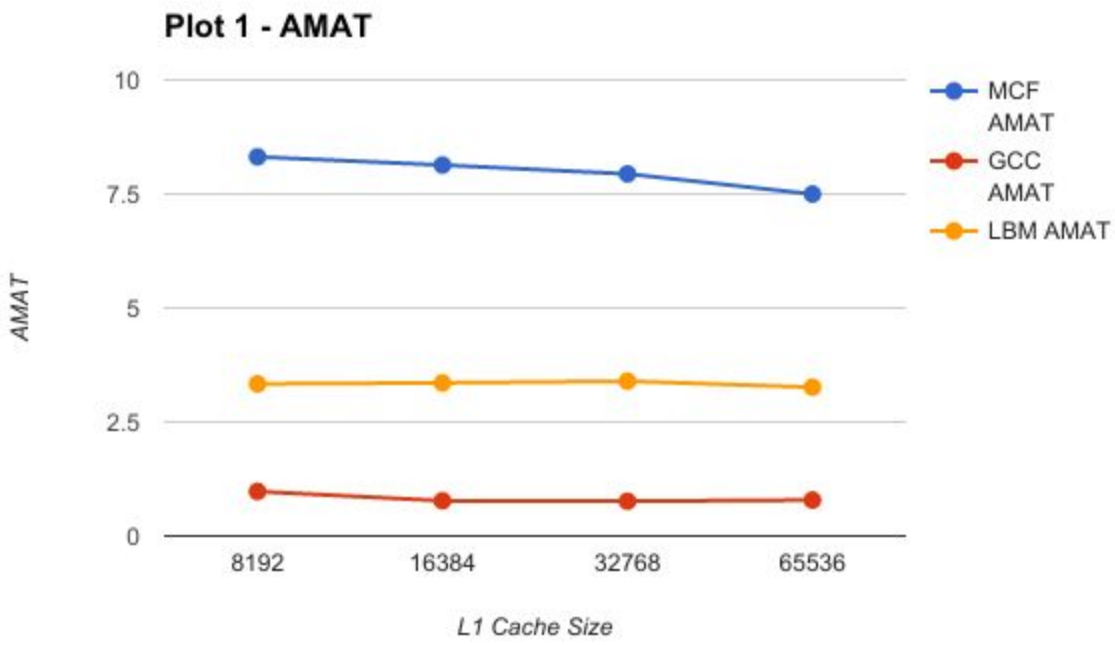
Project 1 Summary:
In this project, I implemented a cache hierarchy simulator. Two-level cache miss rates for different cache configurations were estimated based on simulator inputs. Cache behavior with respect to L1 size, L1 associativity, L2 size, L2 associativity, replacement policy and inclusion policy were all studied during this project. Benchmark AMAT are provided based on three trace files. Plots of several cache configurations with discussions are detailed below.

Plot 1



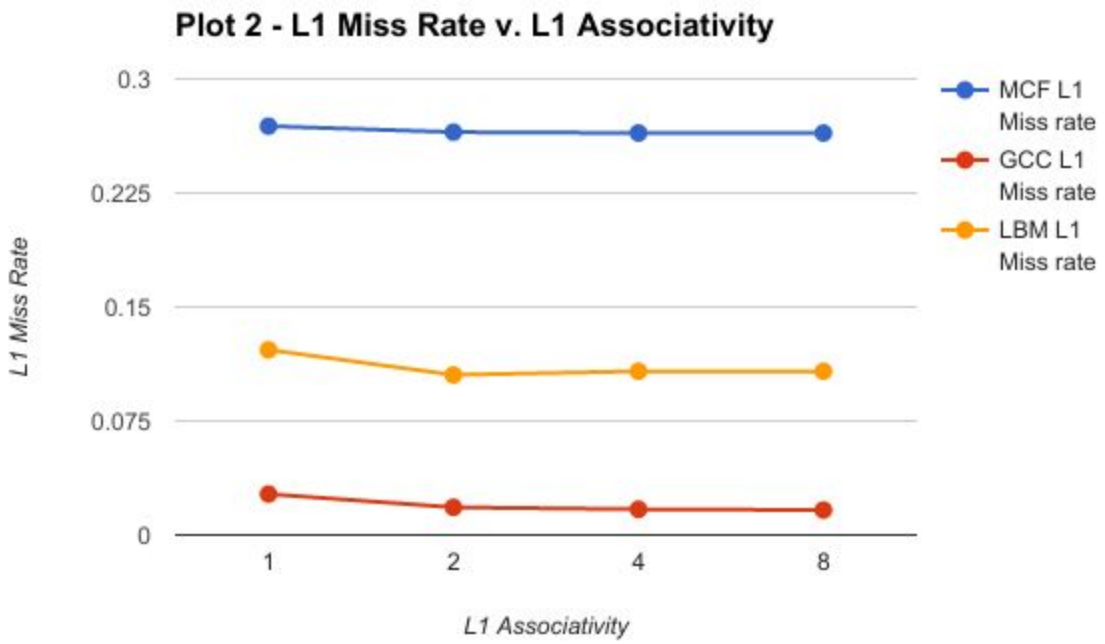
As the L1 cache size increases, the L1 miss rate decreases, but only slightly. Because the L1 miss rate decreases by such small amounts despite large increases in L1 cache size, this signifies a trend of diminishing returns. Each of the three benchmarks above showcase this notion. This would signify that a good design of L1 would be to focus on a shorter hit time rather than having a large L1 cache which is slow.

Plot 1 AMAT



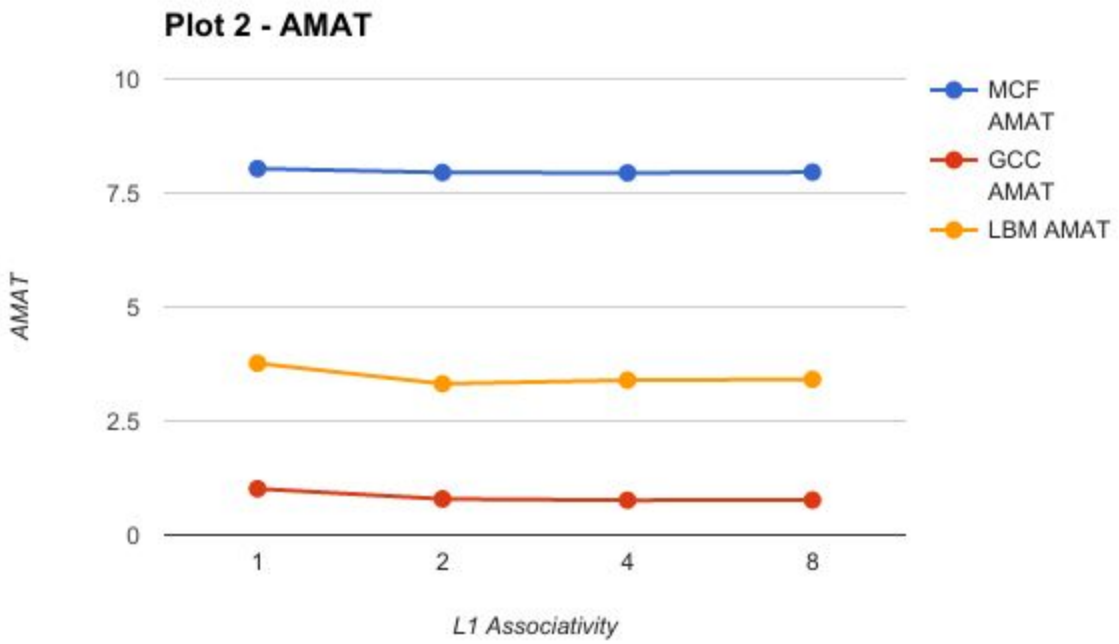
AMAT for the three benchmarks are plotted above with only an L1 cache configuration. The AMAT does not appreciably decrease with increasing L1 cache size. A larger L1 cache would mean higher L1 hit times, but this is not desired in AMAT. Instead, an L1 cache that is fast with a larger L2 cache below would be a good design to implement. L1 cache size differences did not seem to affect the overall AMAT, therefore the focus should be on developing a faster L1.

Plot 2



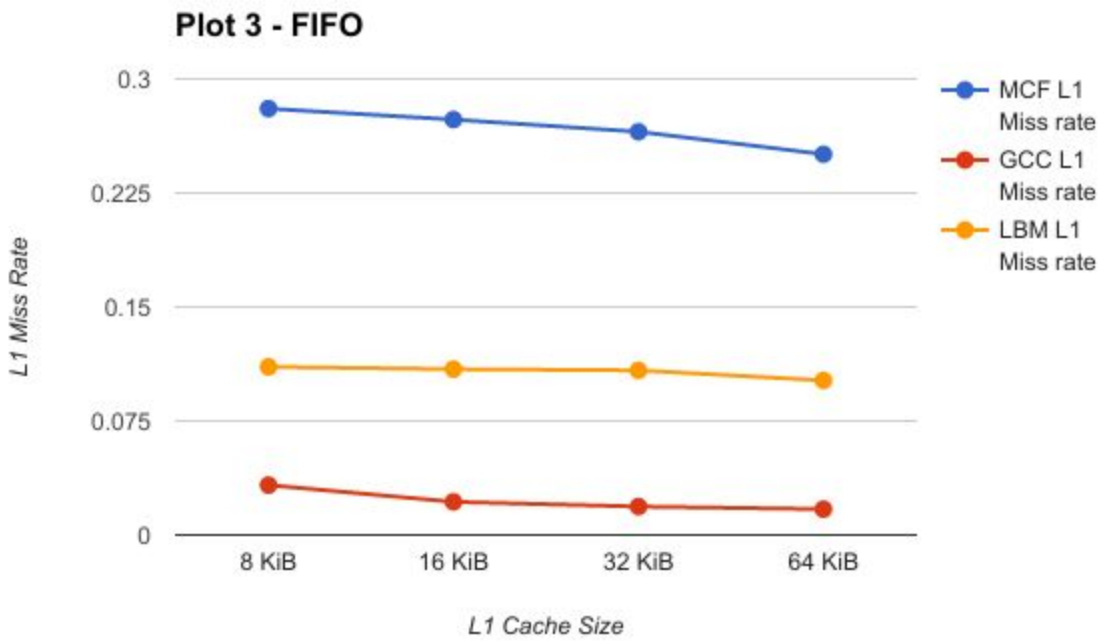
The L1 miss rate versus L1 associativity for only an L1 (no L2) cache configuration is plotted above. Interestingly, while L1 miss rate decreased with greater L1 associativity, 4-way associativity seemed to have the same L1 miss rate as for 8-way associativity. This would suggest that a 4-way or 8-way associative L1 cache would be a good design approximate to a fully associative L1 cache. For the GCC benchmark, the L1 miss rate actually increased as the associativity increased from 2 to 4 (and the 8-way associative L1 miss rate is greater than the 2-way associative L1 miss rate). Increasing L1 associativity would mean longer L1 hit times as there would be more tags per set to compare. This is not a desirable option for improving L1 cache efficiency. Again, the design focus should be on creating a fast L1.

Plot 2 AMAT

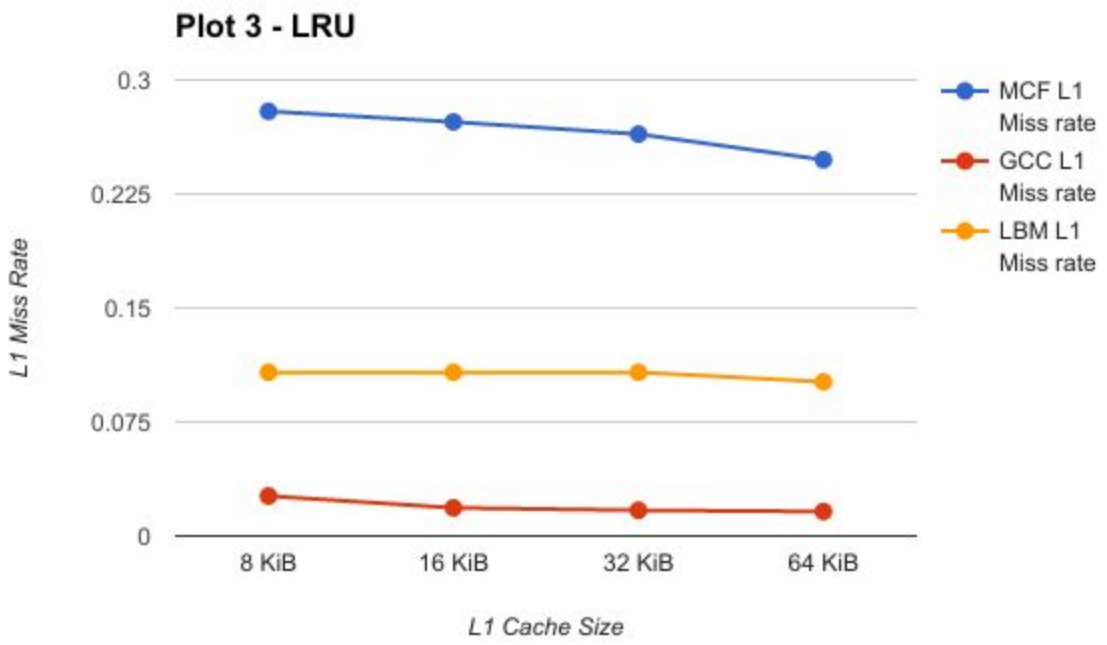


The AMAT plot above shows that increasing L1 associativity does not substantially differ the AMATs. Indeed, for the GCC benchmark, the AMAT actually increases when L1 associativity jumps from 2 to 4. While increasing L1 associativity is good for reducing the L1 miss rate, it does not translate to a better AMAT. A lower L1 miss rate does not necessarily translate to a lower AMAT.

Plot 3



FIFO for an L1 cache is plotted above. Increasing the L1 cache size did tend to decrease the L1 miss rate with the FIFO algorithm. This consistently matches earlier plots of L1 cache size versus L1 miss rate.

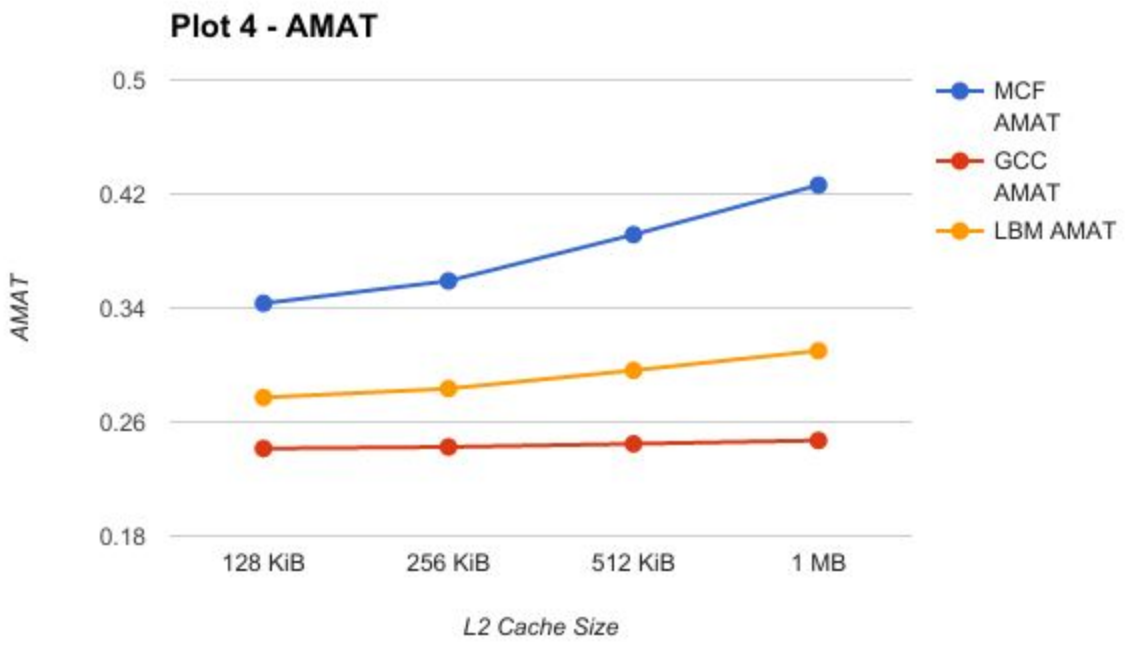


The plot above shows L1 cache size versus L1 miss rate for a configuration that utilizes an LRU replacement policy. As with FIFO, larger L1 caches did not decrease the L1 miss rate significantly. Comparing individual benchmark traces in FIFO with their counterparts in LRU, it can be determined that a better design would implement a FIFO algorithm since it is less complex in hardware while producing approximately the same L1 miss rates for the same L1 cache sizes. A cache that uses a FIFO replacement policy will seem to perform about the same as a cache that implements an LRU policy, so the designer should choose FIFO because it is overall less complex.

Plot 4

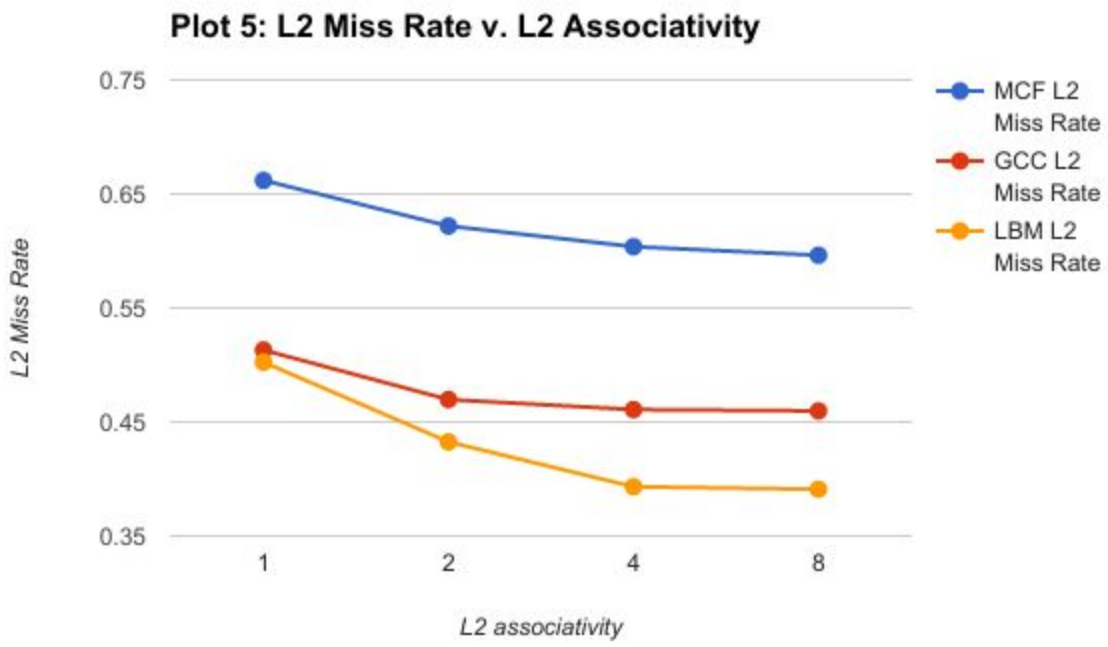
For the MCF benchmark, the L2 miss rate decreased as the L2 cache size increased from 128 KiB to 256 KiB to 512 KiB to 1 MB. For the GCC benchmark, the L2 miss rate stayed relatively consistent despite the increase in L2 cache size. For the LBM benchmark, the L2 miss rate decreased when increasing the L2 cache size from 128 KiB to 256 KiB, but stayed relatively consistent when the L2 cache size was 256 KiB or larger. This illustrated areas of diminishing returns as larger caches tend to increase the hit time because they are slower to access. Other considerations about larger caches include the typically higher power consumption and higher cost.

Plot 4 AMAT



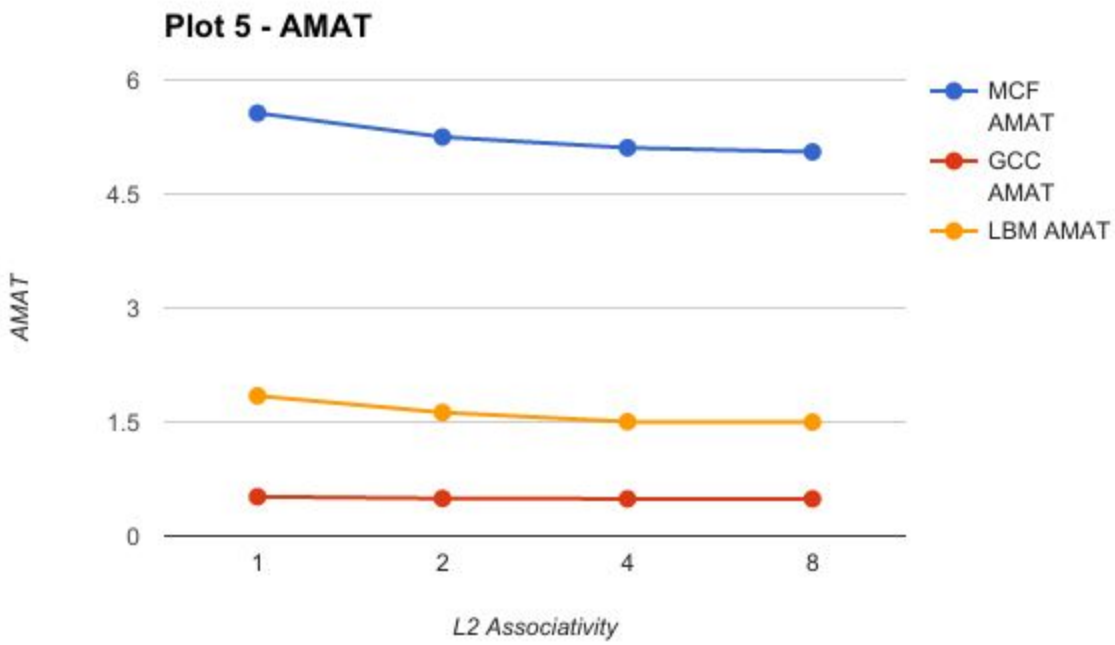
From the AMAT plot above, with increasing L2 cache size, AMAT increased. For the MCF benchmark, this was very noticeable as the L2 cache size grew beyond 256 KiB. For the MCF benchmark, the L2 miss rate decreased with a larger L2 cache size, but its AMAT increased. A lower L2 miss rate does not necessarily translate to a lower AMAT. With a decent amount of associativity, a larger L2 cache may not be desireable as it will increase the L2 hit time, thereby increasing AMAT. Selecting a L2 cache with a size greater than the L1 cache would be a better design.

Plot 5



For the MCF benchmark, the L2 miss rate decreased as the L2 associativity increased. However, with each step up in associativity, the change in L2 miss rate decreases by less and less. This again highlights diminishing returns. For the GCC and LBM benchmarks, the L2 miss rate decreased with increasing L2 associativity, but again the decrease in L2 miss rate was less and less in between L2 associativity. Essentially, the L2 miss rates for all three benchmarks were about the same for both 4-way associativity and 8-way associativity. Speculating from here, a 4-way or 8-way associative cache would behave similarly to a fully associative cache. Considerations for higher associativity include increasing hit time as there are more tags per set to compare, and additional power consumption.

Plot 5 AMAT



From the AMAT plot above, it seems that increasing L2 associativity will slightly increase AMAT. This is apparent for all three benchmarks. Increasing L2 associativity will increase the L2 hit time, as there are more tags per set to match, but the L2 miss rate does considerably drop when increasing associativity from 1 to 2 and 2 to 4. Overall, associativity does not affect AMAT dramatically. While the increasing associativity did improve L2 miss rate, this did not necessarily translate to a lower AMAT. Because 4-way and 8-way associative L2 miss rates were about the same with similar AMAT, a good design would be to use 4-way associativity in lieu of 8-way associativity since the AMAT will be lower without a detrimental tradeoff in L2 miss rate.