

ECEn 528

Study Guide – Advanced caching

- Read Sections 2.1-2.2 and 2.5-2.6 of H&P
 - Things to focus on
 - Advantages and disadvantages of the 10 advanced optimizations. Why they work and when they do not work.
 - Clarifications
 - Two more techniques:
 - If data can be read from a write buffer of a writeback cache, we call it a **victim buffer**
 - If the victim buffer is generalized to also hold clean evicted lines, we call it a **victim cache** -- the effect is to “extend” the associativity of the cache for sets which have high conflicts.
 - Answer the following questions:

1. List the advantages and disadvantages of each optimization:

<i>Optimization</i>	<i>Advantages</i>	<i>Disadvantages</i>
Small and simple		
Way prediction		
Pipelined access		
Nonblocking caches		
Banked caches		
Critical word first/early restart		
Merging write buffers		
Compiler optimizations		
Hardware prefetching		
Compiler-controlled prefetching		

2. Why are “small and simple” designs particularly appropriate for first-level caches?
3. Why is it hard to evaluate the performance of nonblocking caches?

4. For what kinds of cache access patterns would merging write buffers be particularly effective?
5. What happens if you perform blocking with a submatrix size which does not fit in the cache?
What if the submatrix size is only 50% of the cache size?
6. Why might hardware prefetching cause a loss of performance?
7. Why are non-faulting software prefetches generally preferred to faulting software prefetches?
8. Why do most of the scientific codes have lower I\$ misses and higher D\$ misses than most of the integer codes (in Figure 5.20)?