作业(4)：TLP

截至时间：2022.12.26/周一 23:59:59
提交方式：超算习堂（https://easyhpc.net/course/157）

Q1: Assume there is a system with two processors (P0 and P1) sharing memory and each has its own private cache. Caches are organized as direct-mapped, and the coherence protocol is snooping-based MSI. At one moment, the snapshot of caches is as shown by the table below:

	Block #	Tag	Data word 1	Data word 2	Coherency State
P0	0	1000	10	20	M
	1	4000	500	600	S
	...				
	N	3000	2	4	S
P1	0	1000	10	10	I
	1	8000	500	600	S
	...				
	N	3000	2	4	S

a. For either P0 or P1, suppose the cache capacity is 16KB, and each block contains two words (as shown in the table, each word is of 4B), then please split the 32b address into tag-index-offset. Use the address notation *A[N:M]* to describe which address bits are used for each part.

b. If P0 is to update *Block 0*. What will happen?

c. If P1 is to update *Block 1*. Should *Block 1* on P0 be invalidated? Why or why not?

d. If P0 is to update *Block N*. How the coherency state on P0 and P1 will be changed?

e. If P1 brings in a block *M* for reading, and no cache has a copy, what state will *Block M* be in? Why?

f. Suppose the MSI protocol is extended to MESI. Will answer be the same for question e? Why or why not?

Q2: Consider the following code segments running on two processors P1 and P2. Assume A, and B, are initially 0. Assume the processor memory model is sequential consistent (SC).

P1	P2
A = 1; A = 2; while (B == 0);	B = 1; while (A != 1); B = 2;

a. What could be the possible final values of B in the SC processor? List all possible values, and explain by giving the execution order.

b. Explain how a compiler optimization might make it impossible for B to be ever set to 2 in the SC processor.

Q3: [You are REQUIRED to answer in English] Read the paper *HeteroGen: Automatic Synthesis of Heterogeneous Cache Coherence Protocols* (HPCA'2022, Link: https://ieeexplore.ieee.org/abstract/document/9773254), then answer the following questions:

a. What are the challenges on coherence and consistency in designing heterogeneous processors?

b. What is compound consistency model?

c. How does *Heterogen* work? Give the brief steps.

d. What are the usages of *Heterogen*? Give some examples.

e. Is *Hererogen* working well, with regard to performance? Please expand.

f. What you have learnt from the paper? List three points, and briefly describe in several sentences.